# Estimation and Forecasting of Dynamic Conditional Covariance: A Semiparametric Multivariate Model

Xiangdong Long[a], Liangjun Su[b], Aman Ullah[c]

[a] *Judge Business School, University of Cambridge, xl256@cam.ac.uk*

[b] *School of Economics, Singapore Management University, ljsu@smu.edu.sg*

[c] *Department of Economics, University of California, Riverside, aman.ullah@ucr.edu*

First Version: August 2004

This Version: March 2009

**ABSTRACT**

The existing parametric multivariate generalized autoregressive conditional heteroskedasticity (MGARCH) model could hardly capture the nonlinearity and the non-normality, which are widely observed in financial data. We propose semiparametric conditional covariance (SCC) model to capture the information hidden in the standardized residuals and missed by the parametric MGARCH models. Our two-stage SCC estimator incorporates the parametric and nonparametric estimators of the conditional covariance in a multiplicative way. We prove the consistency and asymptotic normality of our semiparametric estimator. We conduct a small set of Monte Carlo experiments to demonstrate the advantage of our SCC estimators over their parametric counterparts in terms of mean squared error. For both in-sample fitting and out-of-sample forecasting conditional covariance matrix, our SCC models also outperform the parametric ones in empirical applications on bivariate stock indices and two stock portfolios with thirty underlying stocks.

**JEL Classifications:** C3; C5; G0

**Key Words:** Conditional Covariance Matrix, Multivariate GARCH, Nadaraya-Watson Estimator, Portfolio, Semiparametric Estimator.

# 1 INTRODUCTION

Since the seminal work of Engle (1982), there has developed a huge literature on modeling the time-varying volatility of economic data in univariate case. Nevertheless, for asset allocation, risk management, hedging and asset pricing, multivariate generalized autoregressive conditional heteroskedasticity (MGARCH) models are of more importance both theoretically and practically because they model the volatility and co-volatility of multiple financial assets jointly. Many recent works have been done in the area of MGARCH models, such as the VECH model of Bollerslev, Engle and Wooldridge (1988), the BEKK model of Baba, Engle, Kraft and Kroner (1991) and Engle and Kroner (1995), the dynamic conditional correlation (DCC) model of Engle (2002) and Engle and Sheppard (2001), the Factor GARCH model of Engle, Ng and Rothschild (1990), to name just a few. However, all these existing MGARCH models including the DCC model share two common features: the normality assumption on the error's distribution and the linearity of dynamic conditional covariance matrix. The exceptions include the regime switching dynamic conditional correlation model of Pelletier (2006), the smooth transition conditional correlation (STCC) model by Silvennoinen and Teräsvirta (2005), and the asymmetric dynamic conditional correlation model by Capiello, Engle and Sheppard (2003), where parametric nonlinear conditional correlation models are used with Gaussian errors, and the copula-based MGARCH model by Lee and Long (2008), where copula is used to construct non-Gaussian errors. The normality assumption is rejected by Fama and French (1993), Richardson and Smith (1993), Longin and Solnik (2001), Ang and Chen (2002), Mashal and Zeevi (2002), and Chen, Fan and Patton (2004). The linear dynamic assumption excludes possible nonlinearity. Once we diverge from linearity, there is too much freedom to specify nonlinearity.

If the parametric model is misspecified in either the joint density function or the functional form of the conditional covariance matrix, parametric estimators of conditional covariance will often be inconsistent. Fortunately, such misspecifications could be avoided by nonparametric estimation techniques because of their ability to capture the unknown nonlinearity. Nevertheless, pure nonparametric estimates are subject to the "curse of dimensionality" and have slow convergence rates.

In this paper, we propose a semiparametric conditional covariance (SCC) model, which combines parametric and nonparametric estimators of conditional covariance matrix in a multiplicative way. We first model the conditional covariance matrix parametrically just like what we do for the conventional parametric MGARCH models. Then we model the conditional covariance of the standardized residual nonparametrically. The estimate of the latter will serve as a nonparametric correction factor for the parametric conditional covariance (PCC) estimator. As surveyed by Mishra, Su and Ullah (2008), a lot of work has been done in the framework of combined estimation: Olkin and Speigelman (1987) in the density function; Glad (1998) and Fan and Ullah (1999) in the conditional mean; Gozalo and Linton (2000) in the conditional heteroskedasticity; and Engle and Gonzalez-Rivera (1991) in the likelihood function, among others. Nevertheless, to the best of our knowledge, there is no combined estimator of conditional covariance matrix.

We provide asymptotic theory for our semiparametric estimator. It possesses several advantages over both pure parametric and nonparametric estimators. First, our SCC model avoids the common shortcomings of parametric MGARCH models on potential misspecifications of functional form and density function. It does not rely on either the distributional assumption on the error term or the parametric functional form of the conditional covariance matrix. Second, when the parametric model is misspecified, the parametric estimator of the conditional covariance is generally inconsistent despite the fact that the finite dimensional parameter in the parametric model may converge to some pseudo-true parameter (see White, 1994). In contrast, our semiparametric estimator can still be consistent with the true conditional covariance matrix under certain conditions. Third, when the parametric model is correctly specified, as expected, our semiparametric estimator is less efficient than the parametric estimator but it can achieve the parametric convergence rate with a fixed bandwidth. Fourth, based on our estimator for the nonparametric correction factor, we propose a test of correct specification of PCC models, which has not been addressed in earlier literature on combined estimation.

We report a small set of Monte Carlo simulation results motivated by the stylized fact that conditional correlation tends to be high during the crisis period and low during the tranquil period. We examine the small sample performance of various PCC and SCC models in terms of mean squared error (MSE). We find that our semiparametric estimators can beat their parametric counterparts in all DGPs under examination. We also apply our strategy to do in-sample (IS) estimating and out-of-sample (OoS) forecasting the conditional covariance matrix of two stock market indices and two stock portfolios with thirty underlying stocks[1]. We consider two types of loss functions, one is the statistical loss function (MSE) and the other is the economic loss function (VaR loss). Again, we find that both of our SCC estimators and forecasters always outperform their parametric counterparts significantly, including the DCC model of Engle (2002).

The rest of the paper is organized as follows. We briefly review some PCC models in Section 2. We present our alternative SCC model and study its asymptotic properties in Section 3. In Section 4 we provide a small set of Monte Carlo experiments and three empirical applications to evaluate the finite sample performance of our SCC models in comparison with some widely used PCC models. We make some concluding remarks in Section 5. All proofs are relegated to Appendix.

To proceed, we define some notation that will be used throughout the paper. Let $\mathbf{I}_k$ denote a $k \times k$ identity matrix. Let $\mathbf{z} = (z_1, ..., z_k)'$ be a $k \times 1$ vector and $\mathbf{Z}$ be a symmetric $k \times k$ matrix with $(i, j)$th element $z_{ij}$. The Euclidean norm of $\mathbf{z}$ or $\mathbf{Z}$ is denoted as $\|\mathbf{z}\|$ or $\|\mathbf{Z}\|$. We define the following

[1]We could apply our method to higher-dimensional portfolio, however, PC's memory card is a constraint.

operators: diag($\mathbf{Z}$) denotes the diagonal matrix with $z_i$ in the $(i,i)$th place; $\mathbf{Z}^*$ denotes a diagonal matrix with the square roots of the diagonal elements of $\mathbf{Z}$ on its diagonal when $\mathbf{Z}$ is positive definite; vec($\mathbf{Z}$) stacks the columns of $\mathbf{Z}$ into a $k^2 \times 1$ vector; vech($\mathbf{Z}$) stacks the lower triangular part of $\mathbf{Z}$ (including the diagonal elements) into a $k\,(k+1)\,/2 \times 1$ vector. Further, we use $D_k$ to denote the $k^2 \times (k\,(k+1)\,/2)$ unique duplication matrix and $D_k^+$ to denote its generalized inverse, which is of size $(k\,(k+1)\,/2) \times k^2$. That is, $\text{vec}(\mathbf{Z}) = D_k \text{vech}(\mathbf{Z})$, $\text{vech}(\mathbf{Z}) = D_k^+ \text{vec}(\mathbf{Z})$, $D_k^+ = (D_k' D_k)^{-1} D_k'$ and $D_k^+ D_k = \mathbf{I}_{k(k+1)/2}$. Here we have used the fact that $D_k' D_k$ is nonsingular. Let $N_k \equiv D_k D_k^+$. We will use the following properties of $N_k$: $N_k$ is symmetric, $N_k D_k = D_k$, $N_k D_k^{+\prime} = D_k^{+\prime}$, and $N_k(\mathbf{A} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{A})N_k$, where $\mathbf{A}$ is a $k \times k$ matrix. For more details, see Magnus and Neudecker (1999, pp. 48-50).

# 2 PARAMETRIC CONDITIONAL COVARIANCE MODELS

Suppose the return series $\{\mathbf{r}_t\}_{t=1}^T$ of the interested financial data follows the stochastic process:

$$\mathbf{r}_t|\mathcal{F}_{t-1} \sim \mathbf{P}(\boldsymbol{\mu}_t, \mathbf{H}_t; \theta),\ t = 1, ..., T, \tag{2.1}$$

where $\mathbf{r}_t \equiv (r_{1,t}, \ldots, r_{k,t})'$ is an $k \times 1$ vector, $\mathcal{F}_{t-1}$ is the information set ($\sigma-$field) at time $t-1$, $E(\mathbf{r}_t|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t$, $E(\mathbf{r}_t\mathbf{r}_t'|\mathcal{F}_{t-1}) = \mathbf{H}_t$, $\mathbf{H}_t$ is the conditional covariance matrix, and $\mathbf{P}$ is the joint cumulative distribution function (CDF) of $\mathbf{r_t}$, and $\theta$ represents the parameters in the distribution. Like Engle (2002), for simplicity we assume the conditional mean $\boldsymbol{\mu}_t$ is zero. If not, necessary standardization should be applied on the data. Thus we can write the model for $\mathbf{r}_t$ as

$$\mathbf{r}_t = \mathbf{H}_t^{1/2}\mathbf{e}_t, \tag{2.2}$$

where $\mathbf{e}_t \equiv \mathbf{H}_t^{-1/2}\mathbf{r}_t$ is the standardized error with $E(\mathbf{e}_t|\mathcal{F}_{t-1}) = \mathbf{0}$ and $E(\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}) = \mathbf{I}_k$. $\mathbf{e}_t$ is typically assumed to follow the standard normal distribution: $\mathbf{e}_t \sim$ i.i.d. $N(0, \mathbf{I}_k)$. We are interested in estimating the conditional covariance matrix $\mathbf{H}_t$ of $\mathbf{r}_t$ without such a distributional assumption.

The conditional covariance matrix $\mathbf{H}_t$ can be decomposed as

$$\mathbf{H}_t = \mathbf{D}_t\,(\theta)\,\mathbf{R}_t\,(\theta)\,\mathbf{D}_t\,(\theta)\,, \tag{2.3}$$

where $\mathbf{R}_t\,(\theta)$ is the conditional correlation matrix with the $(i,j)$th element denoted as $\rho_{ij,t}\,(\theta)$, which stands for the conditional correlation between $r_{i,t}$ and $r_{j,t}$ and can be time-varying; $\mathbf{D_t}\,(\theta)$ =diag($\sqrt{h_{1,t}}$, ..., $\sqrt{h_{k,t}}$) is a diagonal matrix with the square root of the conditional variances $h_{i,t}$, parameterized by the vector $\theta$, on the diagonal. It is well known (see e.g., Engle, 2002) that the conditional correlation matrix $\mathbf{R}_t\,(\boldsymbol{\theta})$ is also the conditional covariance matrix of the standardized returns

$\boldsymbol{\varepsilon}_t \equiv (\varepsilon_{1,t}, \ldots, \varepsilon_{k,t})' = \mathbf{D}_t^{-1}(\theta)\, \mathbf{r}_t$, that is

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' | \mathcal{F}_{t-1}) = \mathbf{R}_t(\theta). \tag{2.4}$$

Now we review some existing parametric models for the conditional covariance matrix $\mathbf{H}_t$, which belong to the class of multivariate GARCH models. These models stem from two different modeling methodologies.

First, both the VECH model by Bollerslev, Engle and Wooldridge (1988) and the BEKK model by Bara, Engle, Kraft and Kroner (1990) and Engle and Kroner (1995) consider modeling the elements of $\mathbf{H}_t$ directly. The VECH model specifies the dynamics of $\mathbf{H}_t$ as

$$\mathrm{vech}(\mathbf{H}_t) = \boldsymbol{\omega} + \Sigma_{i=1}^{p} \mathbf{A}_i \mathrm{vech}(\boldsymbol{\varepsilon}_{t-i} \boldsymbol{\varepsilon}_{t-i}') + \Sigma_{j=1}^{q} \mathbf{B}_j \mathrm{vech}(H_{t-j}), \tag{2.5}$$

where $\omega$ is a $k(k+1)/2 \times 1$ vector, $\mathbf{A}_i$ and $\mathbf{B}_j$ are $k(k+1)/2 \times k(k+1)/2$ matrices. In contrast, the BEKK model specifies $\mathbf{H}_t$ as

$$\mathbf{H}_t = \boldsymbol{\delta}\boldsymbol{\delta}' + \Sigma_{i=1}^{p} \overline{\mathbf{A}}_i \left(\mathbf{r}_{t-i} \mathbf{r}_{t-i}'\right) \overline{\mathbf{A}}_i' + \Sigma_{j=1}^{q} \overline{\mathbf{B}}_j \mathbf{H}_{t-j} \overline{\mathbf{B}}_j', \tag{2.6}$$

where $\boldsymbol{\delta}$ is a $k \times k$ low-triangle matrix, and different matrix properties of $\overline{\mathbf{A}}_i$ and $\overline{\mathbf{B}}_j$ lead to three types of BEKK models: $\overline{\mathbf{A}}_i$ and $\overline{\mathbf{B}}_j$ in full BEKK model, diagonal BEKK model, and scalar BEKK model are full matrix, diagonal matrix, and scalar, respectively.

Second, instead of modeling the conditional covariance matrix directly, some researchers observe $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$ in (2.3) and model $\mathbf{H}_t$ indirectly through modeling $\mathbf{D}_t$ and $\mathbf{R}_t$ separately. The resulting models include the CCC model by Bollerslev (1990), the VC model by Tse and Tsui (2002), the DCC model by Engle (2002) and Engle and Sheppard (2001), among others.

(1) The CCC model assumes that $\mathbf{R}_t = \mathbf{R}$, a constant matrix, and hence the time-varying feature of conditional covariance could only be attributed to the time-varying conditional variances. Nevertheless, Longin and Solnik (2001), Ang and Chen (2002), and Andersen, Bollerslev, Diebold and Labys (1999) indicate asymmetric phenomena in conditional correlation, that is, high correlation tends to be associated with high volatility or crisis period and low correlation tends to be associated with low volatility or tranquil period. The CCC model is thus defaulted and many researchers turn to time-varying conditional correlation, such as the VC and DCC models.

(2) The VC model by Tse and Tsui (2002) specifies univariate GARCH($\overline{p}, \overline{q}$) models for individual return $r_{it}$ :

$$h_{i,t} = \omega_i + \Sigma_{l=1}^{\overline{p}} \kappa_{l,i} h_{i,t-l} + \Sigma_{l=1}^{\overline{q}} \lambda_{l,i} r_{i,t-l}^2, \tag{2.7}$$

and GARCH-type dynamic evolutions for the conditional correlation process $\{\mathbf{R}_t\}$:

$$\mathbf{R}_t = \left(1 - \Sigma_{i=1}^{m} \beta_i - \Sigma_{j=1}^{n} \gamma_j\right) \bar{\mathbf{R}} + \Sigma_{i=1}^{m} \beta_i \mathbf{R}_{t-i} + \Sigma_{j=1}^{n} \gamma_j \widehat{\mathbf{R}}_{t-j}, \tag{2.8}$$

where $\bar{\mathbf{R}}$, $\mathbf{R}_t$, and $\widehat{\mathbf{R}}_t$ are the unconditional, conditional, and sample correlation matrices at time $t$ with unit diagonal elements; the off-diagonal elements $\widehat{\rho}_{ij,t}$ of $\widehat{\mathbf{R}}_t$ is

$$\widehat{\rho}_{ij,t} = \frac{\sum_{l=1}^{M} \varepsilon_{i,t-l}\varepsilon_{j,t-l}}{\sqrt{(\sum_{l=1}^{M} \varepsilon_{i,t-l}^2)(\sum_{l=1}^{M} \varepsilon_{j,t-l}^2)}}; \tag{2.9}$$

the off-diagonal elements of $\bar{\mathbf{R}}$ lie on the interval $(-1, 1)$; and to guarantee the positiveness of $\widehat{\mathbf{R}}_t$, $M$ should not be less than $k$. Conditional covariance inherits the time-varying property from both the conditional variances and the conditional correlations.

(3) Similar to the CCC and VC models, the DCC model by Engle (2002) and Engle and Sheppard (2001) also uses two-stage modeling strategy. In the first stage, one models the conditional variance processes with the usual univariate GARCH models and then obtains the standardized residual $\hat{\boldsymbol{\varepsilon}}_t$. In the second stage, one models the conditional covariance $\mathbf{Q}_t$ of $\boldsymbol{\varepsilon}_t$ as

$$\mathbf{Q}_t = \left(1 - \Sigma_{i=1}^{m}\beta_i - \Sigma_{j=1}^{n}\gamma_j\right) \bar{\mathbf{Q}} + \Sigma_{i=1}^{m}\beta_i(\widehat{\boldsymbol{\varepsilon}}_{t-i}\widehat{\boldsymbol{\varepsilon}}_{t-i}') + \Sigma_{j=1}^{n}\gamma_j\mathbf{Q}_{t-j}, \tag{2.10}$$

where $\bar{\mathbf{Q}}$ is the sample covariance matrix for $\hat{\boldsymbol{\varepsilon}}_t$, $\beta_i > 0$, $\gamma_j > 0$, and $\Sigma_{i=1}^{m}\beta_i + \Sigma_{j=1}^{n}\gamma_j < 1$. The basic properties of correlation matrix, such as positive definiteness and unit diagonal element, are ensured by using the transformation

$$\mathbf{R}_t = \mathbf{Q}_t^{*-1}\mathbf{Q}_t\mathbf{Q}_t^{*-1} \tag{2.11}$$

where $\mathbf{Q}_t^*$ is a diagonal matrix with the square roots of the diagonal elements of $\mathbf{Q}_t$ on its diagonal. Due to its simplicity, the DCC model is flexible for high-dimensional system.

(4) Other specifications for $\mathbf{R}_t$ are also available. For example, Pelletier (2006) develops a Markov-Switching conditional correlation model that allows the conditional correlation to switch between $m$ distinct values by assuming that

$$\mathbf{R}_t = \mathbf{R}_{s_t}, \tag{2.12}$$

where $S_t$ is an unobservable first-order Markov process with $m$ states with $P\left(S_t = j|S_{t-1} = i\right) = p_{ij}$, $i, j = 1, \ldots, m$, as the transition probabilities. However, noting that correlation targeting (substituting unconditional correlation by sample correlation) is not possible in this case, one has to estimate $mk\left(k-1\right)/2 + \left(m-1\right)$ parameters in the second step. Silvennoinen and Teräsvirta (2005) consider STCC model,

$$\begin{aligned} \mathbf{R}_t &= \overline{\mathbf{R}}_1\left(1 - F\left(x_t; \delta, \lambda\right)\right) + \overline{\mathbf{R}}_2 F\left(x_t; \delta, \lambda\right), & (2.13) \\ F\left(x_t; \delta, \lambda\right) &= \left(1 + \exp\left(-\delta\left(x_t - \lambda\right)\right)\right)^{-1}, \text{ say,} & (2.14) \end{aligned}$$

where $\overline{\mathbf{R}}_1$ and $\overline{\mathbf{R}}_2$ are positive definite correlation matrices, $x_t$ is a scalar transition variable that belongs to the information set $\mathcal{F}_{t-1}$ and $\delta$ determines the smoothness of $F\left(\cdot\right)$ as $x_t$ increases. A crucial element of the STCC is the choice of $x_t$, which could be taken as the standardized lagged return or market volatility. Correlation targeting is not possible in this stage either, and the estimation can be carried out iteratively by concentrating the likelihood.

In addition to the above two classes of methodologies, there are some other models, where $\mathbf{H}_t$ is indirectly studied, such as the Orthogonal GARCH (O-GARCH) or principal components GARCH method by Ding (1994), Alexander (1998, 2001), and the Factor GARCH model of Engle, Ng and Rothschild (1990), where some factors driving the economy or the market are considered.

In the existing parametric models given above, the functional form of covariance matrix are assumed to be of known linear or nonlinear form and the maximum likelihood estimation is done under the assumption of normality. In the next section, we present our semiparametric estimation of the conditional covariance matrix which requires neither.

# 3 AN ALTERNATIVE SEMIPARAMETRIC CONDITIONAL COVARIANCE ESTIMATOR

In this section we first introduce briefly the semiparametric estimator of Hafner, Dijk and Franses (2006, HDF hereafter) and propose an alternative semiparametric estimator for conditional covariance matrix.

## 3.1 HDF's Semiparametric Estimator

Motivated by the idea that the conditional correlations depend on exogenous factors such as the market return or volatility, HDF propose the following semiparametric model for $\mathbf{r}_t$ :

$$\mathbf{r}_t = \mathbf{D}_t\left(\theta\right)\boldsymbol{\varepsilon}_t,\ E\left(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}\right) = 0,\ E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}\right) = \mathbf{R}\left(\mathbf{x}_t\right), \tag{3.1}$$

where $\mathbf{D}_t\left(\theta\right)$ is as defined before (after (2.3)), and $\mathbf{x}_t$ is observable at time $t-1$ and $\mathbf{x}_t \in \mathcal{F}_{t-1}$. Assuming that $\theta$ can be estimated by $\widehat{\theta}$ at the parametric $\sqrt{T}$-rate, they define standardized residuals by $\widetilde{\boldsymbol{\varepsilon}}_t \equiv \boldsymbol{\varepsilon}_t(\widehat{\theta}) = \mathbf{D}_t(\widehat{\theta})^{-1}\mathbf{r}_t$. Then they regress $\widetilde{\boldsymbol{\varepsilon}}_t\widetilde{\boldsymbol{\varepsilon}}_t'$ on $\mathbf{x}_t$ nonparametrically to obtain $\widetilde{\mathbf{Q}}\left(\mathbf{x}\right)$, the Nadaraya-Watson kernel estimator of $E\left(\widetilde{\boldsymbol{\varepsilon}}_t\widetilde{\boldsymbol{\varepsilon}}_t'|\mathbf{x}_t = x\right)$. Their semiparametric conditional correlation matrix estimator is defined by

$$\widetilde{\mathbf{R}}\left(\mathbf{x}\right) = (\widetilde{\mathbf{Q}}^*\left(\mathbf{x}\right))^{-1}\widetilde{\mathbf{Q}}\left(\mathbf{x}\right)(\widetilde{\mathbf{Q}}^*\left(\mathbf{x}\right))^{-1}, \tag{3.2}$$

where $\widetilde{\mathbf{Q}}^*\left(\mathbf{x}\right)$ is a diagonal matrix with the square roots of the diagonal elements of $\widetilde{\mathbf{Q}}\left(\mathbf{x}\right)$ on its diagonal. Their semiparametric estimator of $\mathbf{H}_t$ can be written as follows

$$\widetilde{\mathbf{H}}_t = \mathbf{D}_t(\widehat{\theta})\widetilde{\mathbf{R}}\left(\mathbf{x}_t\right)\mathbf{D}_t(\widehat{\theta}). \tag{3.3}$$

Clearly, the HDF's estimators require correct specification of the conditional variance process in order to obtain a final consistent conditional correlation or covariance estimator. This is unsatisfactory since it is extremely hard to know a prior the correct form of the conditional variance process. Below we propose an alternative SCC estimator that can be consistent even if the conditional variance process may be misspecified in the first stage and it requires similar assumption to that in (3.1).

## 3.2 An Alternative Semiparametric Estimator

Motivated by Glad (1998) and Mishra, Su, and Ullah (2008), we propose an alternative SCC estimator, which combines in a multiplicative way the *parametric* conditional covariance estimator from the first stage with the *nonparametric* conditional covariance estimator from the second stage. Essentially, this estimator nonparametrically adjusts the initial PCC estimator.

Let $\{\mathbf{H}_t = E(\mathbf{r}_t\mathbf{r}_t'|\mathcal{F}_{t-1})\}$ be the true time-varying conditional covariance process:

$$\mathbf{r}_t = \mathbf{H}_t^{1/2}\mathbf{e}_t,\ E(\mathbf{e}_t|\mathcal{F}_{t-1}) = 0,\ E(\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}) = \mathbf{I}_k, \tag{3.4}$$

where $\mathbf{H}_t^{1/2}$ is the symmetric square root matrix of $\mathbf{H}_t$. Let $\{\mathbf{H}_{p,t}(\theta)\}$ be a parametrically-specified time-varying conditional covariance process for $\mathbf{r}_t$, where $\theta \in \Theta \subset \mathbb{R}^p$ and $\mathbf{H}_{p,t}(\theta) \in \mathcal{F}_{t-1}$. Analogous to Mishra, Su, and Ullah (2008), our estimation strategy builds on the simple identity

$$\mathbf{H}_t = \mathbf{H}_{p,t}(\theta)^{1/2} E\left[\mathbf{e}_t(\theta)\mathbf{e}_t(\theta)'|\mathcal{F}_{t-1}\right]\mathbf{H}_{p,t}(\theta)^{1/2}, \tag{3.5}$$

where $\mathbf{H}_{p,t}(\theta)^{1/2}$ is the symmetric square root matrix of $\mathbf{H}_{p,t}(\theta)$, and $\mathbf{e}_t(\theta) = \mathbf{H}_{p,t}(\theta)^{-1/2}r_t$ is the standardized error from the parametric model. When $\theta = \theta_*$, some pseudo-true parameter value, we write $\mathbf{H}_{p,t} = \mathbf{H}_{p,t}(\theta_*)$ and $\mathbf{e}_t = \mathbf{e}_t(\theta_*)$. It is clear that the parametric component $\mathbf{H}_{p,t}(\theta)$ in (3.5) can be any PCC model reviewed in Section 2 and estimated by some standard parametric method. To propose a reasonable estimator for the nonparametric component $E\left[\mathbf{e}_t(\theta)\mathbf{e}_t(\theta)'|\mathcal{F}_{t-1}\right]$, we follow the HDF's idea and assume that the conditional expectation of $\mathbf{e}_t\mathbf{e}_t'$ depends on the current information set $\mathcal{F}_{t-1}$ only through a $q \times 1$ observable vector $\mathbf{x}_t = (x_{1t}, ..., x_{qt})'$. That is,

$$E[\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}] = \mathbf{H}_{np}(\mathbf{x}_t), \tag{3.6}$$

where $\mathbf{x}_t \in \mathcal{F}_{t-1}$. There is a fundamental difference between (3.6) and the last expression in (3.1). In order for $\mathbf{R}(\mathbf{x}_t)$ in (3.1) to be a conditional correlation matrix, the conditional variance matrix or equivalently $\{\mathbf{D}_t(\theta)\}$ has to be specified correctly. Fortunately there is no such a requirement for our definition of $\mathbf{H}_{np}(\mathbf{x}_t)$.

Let $\mathbf{H}_{np,t} = \mathbf{H}_{np}(\mathbf{x}_t)$. (3.5) then reduces to

$$\mathbf{H}_t = \mathbf{H}_{p,t}^{1/2}\mathbf{H}_{np,t}\mathbf{H}_{p,t}^{1/2}. \tag{3.7}$$

Based upon (3.5)-(3.7), we can estimate $\mathbf{H}_t$ in two stages:

*Stage 1: Estimate the parameter* $\theta$ *by* $\widehat{\theta}$ *in the parametric specification* $\{\mathbf{H}_{p,t}(\theta)\}$ *for the conditional covariance process. Define the standardized residuals by* $\widehat{\mathbf{e}}_t = \widehat{\mathbf{H}}_{p,t}^{-1/2}\mathbf{r}_t$, *where* $\widehat{\mathbf{H}}_{p,t} = \mathbf{H}_{p,t}(\widehat{\theta})$.

*Stage 2: Estimate* $E[\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}, \mathbf{x}_t = \mathbf{x}]$ *nonparametrically by*

$$\widehat{\mathbf{H}}_{np}(\mathbf{x}) = \frac{\sum_{s=1}^{T}\widehat{\mathbf{e}}_s\widehat{\mathbf{e}}_s' K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})}{\sum_{s=1}^{T} K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})}, \tag{3.8}$$

*where*

$$K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}) = \prod_{l=1}^{q} h_l^{-1} k\left(\frac{x_{ls} - x_l}{h_l}\right)$$

$\mathbf{h} = (h_1, ..., h_q)$, $h_l = h_l(T)$, $l = 1, ..., q$, *are bandwidth parameters, and* $k$ *is a kernel function. Let* $\widehat{\mathbf{H}}_{np,t} = \widehat{\mathbf{H}}_{np}(\mathbf{x}_t)$. *Then our SCC estimator of* $\mathbf{H}_t$ *is obtained as*

$$\widehat{\mathbf{H}}_{sp,t} = \widehat{\mathbf{H}}_{p,t}^{1/2}\widehat{\mathbf{H}}_{np,t}\widehat{\mathbf{H}}_{p,t}^{1/2}. \tag{3.9}$$

*Correspondingly, the estimator of conditional correlation matrix from our SCC model is*

$$\widehat{\mathbf{R}}_{sp,t} = \left(\widehat{\mathbf{H}}_{sp,t}^{*}\right)^{-1}\widehat{\mathbf{H}}_{sp,t}\left(\widehat{\mathbf{H}}_{sp,t}^{*}\right)^{-1}, \tag{3.10}$$

*where* $\widehat{\mathbf{H}}_{sp,t}^{*}$ *is a diagonal matrix with the square roots of the diagonal elements of* $\widehat{\mathbf{H}}_{sp,t}$ *on its diagonal.*

To proceed, we make a few remarks.

**Remark 1.** When $k = 1$, $\widehat{\mathbf{H}}_{sp,t}$ reduces to the semiparametric estimator of conditional variance in the spirit of Mishra, Su and Ullah (2008) who use local polynomial estimation technique instead.

**Remark 2.** When the parametric model $\mathbf{H}_{p,t}$ is correctly specified, i.e., $\mathbf{H}_{p,t}(\theta_0) = \mathbf{H}_t$ for some $\theta_0 \in \Theta$ and $\theta_0 = \theta_*$, we have:

$$\mathbf{H}_{np}(\mathbf{x}_t) = E[\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}] = \mathbf{I}_k. \tag{3.11}$$

In this case, $\widehat{\mathbf{H}}_{np,t}$ is estimating the $k \times k$ identity matrix. On the other hand, if the parametric model $\mathbf{H}_{p,t}$ is misspecified, $\mathbf{H}_{np}(\mathbf{x}_t)$ will not be an identity matrix, and $\widehat{\mathbf{H}}_{np,t}$ will serve as an nonparametric correction factor, which nonparametrically adjusts the initial PCC estimator. In Section 3.4 we will propose a test of correct specification of parametric conditional covariance models based on (3.11).

**Remark 3.** Our SCC estimator is quite different from that of HDF. In the special case where $\widehat{\mathbf{H}}_{p,t}^{1/2} = \mathbf{D}_t(\widehat{\theta})$, then $\widehat{\mathbf{H}}_{np,t}$ is the same as $\widetilde{\mathbf{Q}}(\mathbf{x}_t)$ obtained by HDF. So

$$\widehat{\mathbf{H}}_{sp,t} = \mathbf{D}_t(\widehat{\theta})\widetilde{\mathbf{Q}}(\mathbf{x}_t)\mathbf{D}_t(\widehat{\theta}).$$

We can show that $\widehat{\mathbf{H}}_{sp,t}$ is asymptotically equivalent to $\widetilde{\mathbf{H}}_t = \mathbf{D}_t(\widehat{\theta})(\widetilde{\mathbf{Q}}^*(\mathbf{x}_t))^{-1}\widetilde{\mathbf{Q}}(\mathbf{x}_t)(\widetilde{\mathbf{Q}}^*(\mathbf{x}_t))^{-1}\mathbf{D}_t(\widehat{\theta})$. In the general case where $\widehat{\mathbf{H}}_{p,t}^{1/2} \neq \mathbf{D}_t(\widehat{\theta})$, $\widehat{\mathbf{H}}_{np,t}$ is not equal to $\widetilde{\mathbf{Q}}(\mathbf{x}_t)$ and $\widehat{\mathbf{H}}_{sp,t}$ and $\widetilde{\mathbf{H}}_t$ may have quite different properties in both large and small samples. If the parametric models $(\mathbf{H}_{p,t}(\theta)$ in our case and $\mathbf{D}_t(\theta)$ in HDF's case) are misspecified, our estimator for the conditional covariance matrix is still consistent under weak conditions while that of HDF is generally inconsistent.

**Remark 4.** In the above analysis, we assume $\mathbf{x}_t$ is observable. It turns out this is not necessary. In fact, we can allow $\mathbf{x}_t$ to be estimated from the data at a certain rate. See Mishra, Su, and Ullah (2008).

**Remark 5.** There are some alternatives to obtain the semiparametric estimators. For example, instead of using the Nadaraya-Watson (local constant) estimator, one can obtain the local polynomial estimator (e.g., Fan and Gijbels, 1996; Pagan and Ullah, 1999).

## 3.3 Asymptotic Properties of Our SCC Estimator

To study the asymptotic property of our SCC estimator, we make the following set of assumptions.

**Assumptions**

(A1) The strictly stationary process $\{\mathbf{r}_t, \mathbf{x}_t\}$ is $\alpha$-mixing with mixing coefficients $\alpha(j)$ satisfying $\sum_{j=1}^{\infty} j^a \alpha(j)^{\delta/(\delta+2)} < \infty$ for some $\delta > 0$ and $a > \delta/(\delta+2)$. Also, $E\left(\|\mathbf{r}_t\|^{2(2+\delta)}\right) < \infty$ and $E\left(\|\mathbf{x}_t\|^{2+\delta}\right) < \infty$.

(A2) The pseudo-true parameter $\theta_* \in \Theta \subset \mathbb{R}^p$ governing the PCC process $\{\mathbf{H}_{p,t}(\theta)\}$ exists uniquely and lies in the interior of a compact set $\Theta$.

(A3) $\widehat{\theta} - \theta_* = O_P(T^{-1/2})$.

(A4) $\mathbf{H}_{p,t} = \mathbf{H}_{p,t}(\theta_*)$ is symmetric, finite, and positive definite for each $t$. The process $\{\mathbf{e}_t = \mathbf{H}_{p,t}^{-1/2} r_t\}$ is strictly stationary and $\alpha$-mixing with mixing coefficients $\alpha(j)$. $\mathbf{x}_t$ has a continuous density $f(\mathbf{x})$ that is bounded away from zero at $\mathbf{x}$.

(A5) Let $\mathbf{H}_{p,t}(\theta)$ has continuous derivatives in the neighborhood of $\theta_*$. $\mathbf{H}_{np}(\mathbf{x})$ have two continuous derivatives in the neighborhood of $\mathbf{x}$. For some $\epsilon > 0$, $\sup_{\{\theta:\|\theta-\theta_*\|\le\epsilon\}} \|\boldsymbol{\xi}_t(\theta)\| \le \overline{D}_t$, where $\boldsymbol{\xi}_t(\theta) = \partial \mathbf{e}_t(\theta)/\partial\theta'$ and $E(\overline{D}_t^2) < \infty$.

(A6) Let $\mu_{ij} = \int u^i k(u)^j du$. The kernel $k(\cdot)$ is a symmetric bounded density function such that $\mu_{21} < \infty$ and $|uk(u)| \to 0$ as $|u| \to \infty$.

(A7) As $T \to \infty$, $h_j \to 0$, $T\mathbf{h}! \to \infty$, and $\overline{\lim} T \|\mathbf{h}\|^4 \mathbf{h}! = c \in [0, \infty)$, where $\mathbf{h}! = \Pi_{j=1}^q h_j$.

Assumption A1 is a high-level assumption. When the individual return series follows a GARCH(1,1) process, HDF shows that the $\alpha$-mixing of $\{\mathbf{r}_t\}$ can be satisfied under weak conditions. Assumptions A2-A3 do not require the correct specification for modeling the parametric component. For example, whether the parametric model is true or not, under some regularity conditions for quasi maximum likelihood estimation QMLE, the pseudo true parameter $\theta_*$ exists uniquely (White, 1994, Ch.2) and can be estimated consistently at the regular $\sqrt{T}$ rate (White, 1994, Ch.6). Assumptions 4-5 impose some regularity conditions on the $\{\mathbf{H}_{p,t}(\theta)\}$ process. Assumptions A6-A7 are standard in the nonparametric kernel estimation literature.

The following theorem establishes the asymptotic property of $\widehat{\mathbf{H}}_{np}(\mathbf{x})$.

**Theorem 3.1** *Under Assumptions A1-A7,*

$$\sqrt{T\mathbf{h}!}\left\{vech(\widehat{\mathbf{H}}_{np}(\mathbf{x}))-vech(\mathbf{H}_{np}(\mathbf{x}))-vech(\mathbf{B}(\mathbf{x}))\right\}\xrightarrow{d}N\left(0,\mu_{02}^{q}f(\mathbf{x})^{-1}D_k^{+}\mathbf{\Omega}(\mathbf{x})D_k^{+\prime}\right),\quad(3.12)$$

*where* $\mathbf{\Omega}(\mathbf{x})=(\omega_{ij,lm}(\mathbf{x}))$ *is a* $k^2\times k^2$ *matrix with typical elements*

$$\omega_{ij,lm}(\mathbf{x})=Cov\left(\varrho_{ij,t},\varrho_{lm,t}|\mathbf{x}_t=\mathbf{x}\right)\ \text{with}\ \varrho_{ij,t}=e_{it}e_{jt},$$

$\mathbf{B}(\mathbf{x})=(\mathbf{B}_{ij}(\mathbf{x}))$ *is a* $k\times k$ *matrix with typical elements*

$$\mathbf{B}_{ij}(\mathbf{x})=\frac{\mu_{21}}{2f(\mathbf{x})}\sum_{l=1}^{q}\left[2\frac{\partial f(\mathbf{x})}{\partial x_l}\frac{\partial\mathbf{H}_{np,ij}(\mathbf{x})}{\partial x_l}+f(\mathbf{x})\frac{\partial^2\mathbf{H}_{np,ij}(\mathbf{x})}{\partial x_l\partial x_l}\right]h_l^2,$$

*where* $e_{it}$ *is the ith element of* $\mathbf{e}_t$ *and* $\mathbf{H}_{np,ij}(\mathbf{x})$ *is the* $(i,j)$*th element of* $\mathbf{H}_{np}(\mathbf{x})$.

**Remark 6.** Theorem 3.1 implies that we can estimate $\mathbf{H}_{np}(\mathbf{x})$ consistently by $\widehat{\mathbf{H}}_{np}(\mathbf{x})$, which has the usual asymptotic bias and variance structure as typical local constant estimators. Let $\boldsymbol{\eta}_t$ =vech($\mathbf{e}_t\mathbf{e}_t'$). We can get an alternative expression for $D_k^{+}\mathbf{\Omega}(\mathbf{x})D_k^{+\prime}$:

$$D_k^{+}\mathbf{\Omega}(\mathbf{x})D_k^{+\prime}=\mathrm{Var}(\boldsymbol{\eta}_t|\mathbf{x}_t=\mathbf{x}).$$

When the start-up PCC model is correctly specified, i.e., $\mathbf{H}_t=\mathbf{H}_{p,t}(\theta_*)$, we have: $\mathbf{H}_{np}(\mathbf{x})=\mathbf{I}_k$, the asymptotic bias term in (3.12) vanishes ($\mathbf{B}(\mathbf{x})=0$).

The asymptotic property of our semiparametric estimator for the conditional covariance matrix $\mathbf{H}_t$ is stated in the following corollary.

**Corollary 3.2** *(i) For any* $\mathbf{x}_t$ *such that* $f(\mathbf{x}_t)$ *is bounded away from 0,* $\widehat{\mathbf{H}}_{sp,t}$ *and* $\widehat{\mathbf{R}}_{sp,t}$ *are consistent for* $\mathbf{H}_t$ *and* $\mathbf{R}_t$*, respectively. That is,*

$$\widehat{\mathbf{H}}_{sp,t}=\widehat{\mathbf{H}}_{p,t}^{1/2}\widehat{\mathbf{H}}_{np,t}\widehat{\mathbf{H}}_{p,t}^{1/2}\xrightarrow{p}\mathbf{H}_t,\ \text{and}\ \widehat{\mathbf{R}}_{sp,t}=\left(\widehat{\mathbf{H}}_{sp,t}^{*}\right)^{-1}\widehat{\mathbf{H}}_{sp,t}\left(\widehat{\mathbf{H}}_{sp,t}^{*}\right)^{-1}\xrightarrow{p}\mathbf{R}_t.$$

*(ii)* $\sqrt{T\mathbf{h}!}\left\{vech\left(\widehat{\mathbf{H}}_{sp,t}\right)-vech(\mathbf{H}_t)-\overline{\mathbf{B}}_t(\mathbf{x}_t)\right\}\xrightarrow{d}MN\left(0,\mu_{02}^{q}f(\mathbf{x}_t)^{-1}D_k^{+}\overline{\mathbf{\Omega}}_t(\mathbf{x}_t)D_k^{+\prime}\right),$ *where* $\overline{\mathbf{B}}_t(\mathbf{x})$ $=vech\left(\mathbf{H}_{p,t}^{1/2}\mathbf{B}(\mathbf{x})\mathbf{H}_{p,t}^{1/2}\right)$ *and* $\overline{\mathbf{\Omega}}_t(\mathbf{x})=\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)\mathbf{\Omega}(\mathbf{x})\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)$. *That is, conditional on* $\mathbf{H}_{p,t}$ *and* $\mathbf{x}_t$, $\sqrt{T\mathbf{h}!}\left\{vech\left(\widehat{\mathbf{H}}_{sp,t}\right)-vech(\mathbf{H}_t)-\overline{\mathbf{B}}_t(\mathbf{x}_t)\right\}$ *is asymptotically normal with mean zero and variance* $\mu_{02}^{q}f(\mathbf{x}_t)^{-1}\overline{\mathbf{\Omega}}_t(\mathbf{x}_t)$.

**Remark 7.** Corollary 3.2(i) says that we can obtain a consistent estimator for the conditional covariance and correlation matrix. Corollary 3.2(ii) essentially says that $\widehat{\mathbf{H}}_{sp,t}$ is also asymptotically normally distributed conditional on $\mathbf{H}_{p,t}$ and $\mathbf{x}_t$, and it inherits the asymptotic bias and variance structure of $\widehat{\mathbf{H}}_{np}(\mathbf{x}_t)$. By the delta method, one can also show that the semiparametric estimator for conditional correlation matrix is also asymptotically distributed with the nonparametric convergence rate $\sqrt{T\mathbf{h}!}$.

**Remark 8.** To compare our estimator with the parametric estimator of conditional covariance, first note that when the parametric component is correctly specified, as expected, our estimator is less efficient than the parametric one since our estimator has a slower convergence rate than the parametric estimator as $\|\mathbf{h}\| \to 0$. Nevertheless, when $\mathbf{h}$ is kept fixed, a careful examination of the proof of Theorem 3.1 and Corollary 3.2 indicates that our semiparametric estimator is consistent with the true conditional covariance with the regular parametric $\sqrt{T}$ convergence rate. In this sense, we say that our estimator is almost as good as the parametric estimator in terms of convergence rate when $\mathbf{h}$ is kept fixed. Next, in case of misspecification, the PCC estimator is usually inconsistent (even though $\widehat{\theta}$ is consistent for some pseudo-true parameter $\theta_*$) while our semiparametric conditional covariance estimator is still consistent. Similar remarks hold true for the estimators of conditional correlation matrix.

**Remark 9.** Like, Mishra, Su and Ullah (2008), we can also compare our semiparametric estimator of conditional covariance with the one-step nonparametric kernel estimator. For the ease of comparison, we consider the simplest case where both $\mathbf{H}_{p,t}$ and $\mathbf{H}_t$ depend on the information set $\mathcal{F}_{t-1}$ only through $\mathbf{x}_t$. In this case, we can write $\mathbf{H}_{p,t} = \mathbf{H}_p(\mathbf{x}_t)$ and $\mathbf{H}_t = \mathbf{H}(\mathbf{x}_t)$, and the nonparametric kernel estimator of $\mathbf{H}_t = \mathbf{H}(\mathbf{x}_t)$ is given by

$$\widetilde{\mathbf{H}}_{ncc,t} = \frac{\sum_{s=1}^{T} \mathbf{r}_s \mathbf{r}_s' K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}_t)}{\sum_{s=1}^{T} K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}_t)}.$$

In the sequel, we will call $\widetilde{\mathbf{H}}_{ncc,t}$ as the nonparametric conditional covariance (NCC) estimator. Standard nonparametric regression theory reveals that

$$\sqrt{T\mathbf{h}!}\left\{\text{vech}\left(\widetilde{\mathbf{H}}_{ncc,t}\right) - \text{vech}\left(\mathbf{H}_t\right) - \text{vech}(\mathbf{B}_{ncc}\left(\mathbf{x}_t\right))\right\} \xrightarrow{d} MN\left(0, \mu_{02}^q f(\mathbf{x}_t)^{-1} D_k^+ \mathbf{\Omega}_{ncc}\left(\mathbf{x}_t\right) D_k^{+\prime}\right),$$

where $\mathbf{\Omega}_{ncc}\left(\mathbf{x}\right) = \left(\omega_{ij,lm}^{(ncc)}\left(\mathbf{x}\right)\right)$ is a $k^2 \times k^2$ matrix with typical elements

$$\omega_{ij,lm}^{(ncc)}\left(\mathbf{x}\right) = \text{Cov}\left(r_{it}r_{jt}, r_{lt}r_{mt} | \mathbf{x}_t = \mathbf{x}\right),$$

and $\mathbf{B}_{ncc}\left(\mathbf{x}\right) = \left(\mathbf{B}_{ncc,ij}\left(\mathbf{x}\right)\right)$ is a $k \times k$ matrix with typical elements

$$\mathbf{B}_{ncc,ij}\left(\mathbf{x}\right) = \frac{\mu_{21}}{2f\left(\mathbf{x}\right)} \sum_{l=1}^{q} \left[2\frac{\partial f\left(\mathbf{x}\right)}{\partial x_l}\frac{\partial \mathbf{H}_{ij}\left(\mathbf{x}\right)}{\partial x_l} + f\left(\mathbf{x}\right)\frac{\partial^2 \mathbf{H}_{ij}\left(\mathbf{x}\right)}{\partial x_l \partial x_l}\right] h_l^2, \tag{3.13}$$

where $\mathbf{H}_{ij}\left(\mathbf{x}\right)$ denotes the $(i,j)$th element of $\mathbf{H}\left(\mathbf{x}\right)$, and $r_{it}$ is the $i$th element of $\mathbf{r}_t$.

On the other hand, when both $\mathbf{H}_{p,t}$ and $\mathbf{H}_t$ depend on the information set $\mathcal{F}_{t-1}$ only through

$\mathbf{x}_t$, it is easy to verify that

$$\begin{aligned}
\overline{\mathbf{\Omega}}_t(\mathbf{x}_t) &= \left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)\mathbf{\Omega}(\mathbf{x}_t)\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right) \\
&= \left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)E\left(\text{vec}(\mathbf{e}_t\mathbf{e}_t')\left[\text{vec}(\mathbf{e}_t\mathbf{e}_t')\right]'|\mathbf{x}_t\right)\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right) \\
&= E\left(\text{vec}(\mathbf{H}_{p,t}^{1/2}\mathbf{e}_t\mathbf{e}_t'\mathbf{H}_{p,t}^{1/2})\left[\text{vec}(\mathbf{H}_{p,t}^{1/2}\mathbf{e}_t\mathbf{e}_t'\mathbf{H}_{p,t}^{1/2})\right]'|\mathbf{x}_t\right) \\
&= E\left(\text{vec}(\mathbf{r}_t\mathbf{r}_t')\left[\text{vec}(\mathbf{r}_t\mathbf{r}_t')\right]'|\mathbf{x}_t\right) = \mathbf{\Omega}_{ncc}(\mathbf{x}_t)
\end{aligned}$$

by the fact that $(\mathbf{A}\otimes\mathbf{A})\text{vec}(\mathbf{e}_t\mathbf{e}_t')$ $=\text{vec}(\mathbf{A}\mathbf{e_t}\mathbf{e_t'}\mathbf{A})$ for any $k\times k$ matrix $\mathbf{A}$. This implies that our SCC estimator shares the same asymptotic variance-covariance matrix as the NCC estimator. So we are left to compare the asymptotic bias of our SCC estimator with that of the NCC estimator, i.e., to compare $\overline{\mathbf{B}}_t(\mathbf{x}_t)$ $=\text{vech}(\mathbf{H}_{p,t}^{1/2}\mathbf{B}(\mathbf{x}_t)\mathbf{H}_{p,t}^{1/2})$ with $\text{vech}(\mathbf{B}_{ncc}(\mathbf{x}_t))$.

A typical element of $\overline{\mathbf{B}}_t(\mathbf{x}_t)$ is given by

$$\overline{\mathbf{B}}_{t,ij}(\mathbf{x}_t) = \frac{\mu_{21}}{2f(\mathbf{x})}\sum_{l=1}^{k}\sum_{m=1}^{k}\mathbf{H}_{p,il}^{1/2}(\mathbf{x}_t)\sum_{s=1}^{q}\left[2\frac{\partial f(\mathbf{x}_t)}{\partial x_s}\frac{\partial\mathbf{H}_{np,lm}(\mathbf{x}_t)}{\partial x_s}+f(\mathbf{x})\frac{\partial^2\mathbf{H}_{np,lm}(\mathbf{x}_t)}{\partial x_s\partial x_s}\right]h_s^2\mathbf{H}_{p,mj}^{1/2}(\mathbf{x}_t) \tag{3.14}$$

where $\mathbf{H}_{p,il}^{1/2}(\mathbf{x})$ denotes the $(i,l)$th element of $\mathbf{H}_p^{1/2}(\mathbf{x})$ and $\mathbf{H}_{np,lm}(\mathbf{x})$ is similarly defined. Unfortunately, the above expression generally appears too complicated to compare with $\mathbf{B}_{ncc,ij}(\mathbf{x}_t)$ defined by (3.13). Only in the special case where $k=1$ and $q=1$ and where the local constant method is replaced by the local linear method, can we follow Mishra, Su and Ullah (2008) and show that $\overline{\mathbf{B}}_{t,ij}(\mathbf{x}_t)$ is smaller than $\mathbf{B}_{np,ij}(\mathbf{x}_t)$ in absolute value under weak conditions.

## 3.4 Test the Correct Specification of Parametric Conditional Covariance Model

In this subsection we propose a test of correct specification of parametric conditional covariance models based on (3.11). The null hypothesis is

$$H_0:\mathbf{H}_{np}(\mathbf{x}_t)=\mathbf{I}_k \text{ a.s.} \tag{3.15}$$

and the alternative hypothesis is

$$H_1:\Pr(\mathbf{H}_{np}(\mathbf{x}_t)=\mathbf{I}_k)<1. \tag{3.16}$$

Let $\sigma_{ij}(\mathbf{x})$ denote the $(i,j)$ element of $\mathbf{H}_{np}(\mathbf{x})$, $i,j=1,\cdots,k$. That is, $\sigma_{ij}(\mathbf{x}_t)=E\left[e_{it}e_{jt}|\mathcal{F}_{t-1}\right]$, where recall $e_{it}$ denotes the $i$th element of $\mathbf{e}_t$. We can rewrite the null hypothesis as

$$H_0:P(\sigma_{ij}(\mathbf{x})=\delta_{ij})=1 \text{ for all } i,\ j=1,\cdots,k, \tag{3.17}$$

and the alternative hypothesis as

$$H_1 : \Pr\left(\sigma_{ij}\left(\mathbf{x}\right) = \delta_{ij}\right) < 1 \text{ for some } i,\ j = 1, \cdots, k, \tag{3.18}$$

where $\delta_{ij}$ is Kronecker's delta, i.e., $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Recall that $f\left(\mathbf{x}\right)$ denotes the density function of $\mathbf{x}_t$. When the null and alternative hypotheses are written in the form of (3.17) and (3.18), we can construct consistent tests of $H_0$ versus $H_1$ using various distance measures. A convenient choice is to use the measure

$$\Gamma = \sum_{i=1}^{k-1}\sum_{j=i}^{k}\int\left(\sigma_{ij}\left(\mathbf{x}\right) - \delta_{ij}\right)^2 f^2\left(\mathbf{x}\right)d\mathbf{x} \geq 0 \tag{3.19}$$

and $\Gamma = 0$ if and only if $H_0$ given by (3.17) holds. Note that the use of density weight in the definition of $\Gamma$ will help us avoid the random denominator issue. We will propose a test statistic based upon a kernel estimator of $\Gamma$.

To construct the sample analog of $\Gamma$, we first obtain estimators of $\sigma_{ij}\left(\mathbf{x}\right)$ and $f\left(\mathbf{x}\right)$, which are given by

$$\widehat{\sigma}_{ij}\left(\mathbf{x}\right) = \frac{T^{-1}\sum_{s=1}^{T}\widehat{e}_{it}\widehat{e}_{jt}K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})}{\widehat{f}\left(\mathbf{x}\right)}, \text{ and } \widehat{f}\left(\mathbf{x}\right) = T^{-1}\sum_{s=1}^{T}K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}), \tag{3.20}$$

where $\widehat{e}_{it}$ is the $i$th element of $\widehat{\mathbf{e}}_t$. Note that $\widehat{\sigma}_{ij}\left(\mathbf{x}\right)$ is the $(i, j)$ element of $\widehat{\mathbf{H}}_{np}\left(\mathbf{x}_t\right)$. We then estimate $\Gamma$ by the following functional:

$$\begin{aligned}\widehat{\Gamma}_1 &= \sum_{i=1}^{k-1}\sum_{j=i}^{k}\int\left(\widehat{\sigma}_{ij}\left(\mathbf{x}\right) - \delta_{ij}\right)^2\widehat{f}^2\left(\mathbf{x}\right)d\mathbf{x} \\ &= \frac{1}{T^2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\widehat{e}_{is}\widehat{e}_{js} - \delta_{ij}\right)\left(\widehat{e}_{it}\widehat{e}_{jt} - \delta_{ij}\right)\overline{K}_{\mathbf{h}}\left(\mathbf{x}_s - \mathbf{x}_t\right) \end{aligned}\tag{3.21}$$

where $\overline{K}_{\mathbf{h}}\left(\mathbf{u}\right) = \Pi_{l=1}^{q}h_l^{-1}\overline{k}\left(u_l/h_l\right)$, $\mathbf{u} = (u_1, \cdots, u_q)$, and $\overline{k}\left(u\right) = \int k\left(v\right)k\left(u - v\right)dv$ is the convolution kernel derived from $k$. For example, if $k\left(u\right) = \exp\left(-u^2/2\right)/\sqrt{2\pi}$, then $\overline{k}\left(u\right) = \exp\left(-u^2/4\right)/\sqrt{4\pi}$, a normal density with zero mean and variance 2.

The above statistic is simple to compute and offers a natural way to test $H_0$ in (3.17). Nevertheless, we propose a bias-adjusted test statistic, namely,

$$\widehat{\Gamma} = \frac{1}{T^2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t\neq s}^{T}\left(\widehat{e}_{is}\widehat{e}_{js} - \delta_{ij}\right)\left(\widehat{e}_{it}\widehat{e}_{jt} - \delta_{ij}\right)\overline{K}_{\mathbf{h}}\left(\mathbf{x}_s - \mathbf{x}_t\right). \tag{3.22}$$

In effect, $\widehat{\Gamma}$ removes the "diagonal" ($s = t$) terms from $\widehat{\Gamma}_1$ in (3.21), thus reducing the bias of the statistic. A similar idea has been used in Lavergne and Vuong (2000), Su and White (2007), and Su and Ullah (2008). We will show that after being appropriately scaled, $\widehat{\Gamma}$ is asymptotically normally distributed under suitable assumptions.

To derive the asymptotic properties of the test statistic $\widehat{\Gamma}$, we need to make some additional assumptions.

**Assumptions**

(A8) Let $\varepsilon_{ijt} \equiv e_{it}e_{jt} - \delta_{ij}$. For $i, j = 1, \cdots, k$, $E\left(|\varepsilon_{ijt}|^{4(1+\delta)}\right) \leq C$ and $E\left|\varepsilon_{ijt_1}^{r_1}\varepsilon_{ijt_2}^{r_2}\cdots\varepsilon_{ijt_l}^{r_l}\right|^{1+\delta}$ $\leq C$ for some $C < \infty$, where $2 \leq l \leq 4$, $0 \leq r_s \leq 4$, and $\sum_{s=1}^{l} r_s \leq 8$.

(A9) (i) Let $\mu_{ij2}(\mathbf{x}) \equiv E(\varepsilon_{ijt}^2|\mathbf{x}_t = \mathbf{x})$ and $\mu_{ij4}(\mathbf{x}) = E(\varepsilon_{ijt}^4|\mathbf{x}_t = \mathbf{x})$. Both $\mu_{ij2}(\mathbf{x})$ and $\mu_{ij4}(\mathbf{x})$ satisfy the Lipschitz condition: for $i, j = 1, \cdots, k$ and $l = 2, 4$, $|\mu_{ijl}(\mathbf{x} + \mathbf{x}^*) - \mu_{ijl}(\mathbf{x})| \leq d_{ijl}(\mathbf{x}) \|\mathbf{x}^*\|$, where $\|\cdot\|$ denotes the Euclidean norm and $\int d_{ijl}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} < C < \infty$. (ii) The joint density $f_{t_1,...,t_l}(\cdot)$ of $(\mathbf{x}_{i_1}, ..., \mathbf{x}_{i_l})$ $(1 \leq l \leq 4)$ exists and satisfies the Lipschitz condition: $|f_{t_1,...,t_l}\left(\mathbf{x}^{(1)} + \mathbf{v}^{(1)}, ..., \mathbf{x}^{(l)} + \mathbf{v}^{(l)}\right) - f_{t_1,...,t_l}\left(\mathbf{x}^{(1)}, ..., \mathbf{x}^{(l)}\right)| \leq D_{t_1,...,t_l}\left(\mathbf{x}^{(1)}, ..., \mathbf{x}^{(l)}\right) \|\mathbf{v}\|$, where $\mathbf{v}' = (\mathbf{v}^{(1)'}, ..., \mathbf{v}^{(l)'})$, $\int D_{t_1,...,t_l}\left(\mathbf{v}^{(1)}, ..., \mathbf{v}^{(l)}\right) \|\mathbf{v}\|^{2(1+\delta)} d\mathbf{v} \leq C$ and $\int D_{t_1,...,t_l}\left(\mathbf{v}^{(1)}, ..., \mathbf{v}^{(l)}\right) f_{t_1,...,t_l}\left(\mathbf{v}^{(1)}, ..., \mathbf{v}^{(l)}\right) d\mathbf{v} \leq C$ for some $C < \infty$.

Assumptions A8-A9 are common in nonparametric estimation with strong mixing data (see Gao and King, 2003). They are mainly used in the proof of Theorem 3.3 below.

Define

$$\sigma_0^2 \equiv 2\int \overline{K}^2(\mathbf{u})\, d\mathbf{u} \sum_{i_1=1}^{k-1}\sum_{j_1=i_1}^{k}\sum_{i_2=1}^{k-1}\sum_{j_2=i_2}^{k} E[b_{i_1j_1i_2j_2}^2(\mathbf{x}_t) f(\mathbf{x}_t)],$$

where $b_{i_1j_1i_2j_2}(\mathbf{x}) = E\left[(e_{i_1t}e_{j_1t} - \delta_{i_1j_1})(e_{i_2t}e_{j_2t} - \delta_{i_2j_2})|\mathbf{x}_t = \mathbf{x}\right]$, and $\overline{K}(\mathbf{u}) = \Pi_{l=1}^{q}\overline{k}(u_l)$. The asymptotic null distribution of $\widehat{\Gamma}$ is established in the next theorem.

**Theorem 3.3** *Under Assumptions A1-A9 and under $H_0$, $T(\mathbf{h}!)^{1/2}\widehat{\Gamma} \xrightarrow{d} N(0, \sigma_0^2)$.*

The proof is tedious and is relegated to the appendix. From the proof we know that $T(\mathbf{h}!)^{1/2}\widehat{\Gamma} = T(\mathbf{h}!)^{1/2}\overline{\Gamma} + o_P(1)$, where

$$\overline{\Gamma} = \frac{1}{T^2}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{s=1}^{T}\sum_{t\neq s}^{T}(e_{is}e_{js} - \delta_{ij})(e_{it}e_{jt} - \delta_{ij})\overline{K}_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x}_t)$$

This means that the first stage parametric estimation of the conditional covariance matrix does not affect the first order asymptotic properties of the test. To implement the test, we require a consistent estimate of the variance $\sigma_0^2$. Define

$$\widehat{\sigma}^2 \equiv 2T^{-2}\mathbf{h}!\sum_{s=1}^{T}\sum_{t\neq s}^{T}\left[\sum_{i=1}^{k-1}\sum_{j=i}^{k}(\widehat{e}_{it}\widehat{e}_{jt} - \delta_{ij})(\widehat{e}_{is}\widehat{e}_{js} - \delta_{ij})\right]^2 \overline{K}_{\mathbf{h}}^2(\mathbf{x}_t - \mathbf{x}_s). \tag{3.23}$$

It is easy to show that $\widehat{\sigma}^2$ is consistent for $\sigma_0^2$ under $H_0$. We then compare

$$\widehat{T} \equiv T(\mathbf{h}!)^{1/2}\widehat{\Gamma}/\sqrt{\widehat{\sigma}^2} \tag{3.24}$$

with the one-sided critical value $z_\alpha$ from the standard normal distribution, and reject the null when $\widehat{T} > z_\alpha$.

To examine the asymptotic local power of our test, we consider the following local alternatives:

$$H_1(\gamma_T) : \sigma_{ij}(\mathbf{x}) = \delta_{ij} + \gamma_T \Delta_{ij}(\mathbf{x}),\ i,\ j = 1, \cdots, k, \tag{3.25}$$

where $\Delta_{ij}(\mathbf{x})$ satisfies $E\left|\triangle_{ij}(\mathbf{x}_t)\right|^{2+\delta} < \infty$ and $\gamma_T \to 0$ as $T \to \infty$. Define

$$\Delta_0 \equiv \int \sum_{i=1}^{k-1} \sum_{j=i}^{k} \triangle_{ij}^2(\mathbf{x}) f^2(\mathbf{x}) d\mathbf{x}. \tag{3.26}$$

The following proposition shows that our test can distinguish local alternatives $H_1(\gamma_T)$ at the rate $\gamma_T = T^{-1/2}(\mathbf{h}!)^{-1/4}$.

**Theorem 3.4** *Under Assumptions A.1–A.9, suppose that* $\gamma_T = T^{-1/2}(\mathbf{h}!)^{-1/4}$ *in* $H_1(\gamma_T)$*. Then, the power of the test satisfies* $P(\widehat{T} \geq z_\alpha | H_1(\gamma_T)) \to 1 - \Phi(z_\alpha - \Delta_0/\sigma_0)$*, where* $\Phi(\cdot)$ *is the cumulative distribution function of standard normal.*

Theorem 3.4 implies that the test has non-trivial asymptotic power against alternatives for which $\Delta_0 > 0$. The power increases with the magnitude of $\Delta_0/\sigma_0$. Furthermore, by taking a large bandwidth we can make the alternative magnitude against which the test has non-trivial power, i.e., $\gamma_T$, arbitrarily close to the parametric rate $T^{-1/2}$.

# 4 SIMULATION AND EMPIRICAL ANALYSIS

For now, we show the outperformance of SCC model for IS estimation and OoS predictions via simulation and empirical analysis in the next two subsections.

## 4.1 Monte Carlo Simulation

In this subsection, we follow the idea of Engle (2002) to compare the small sample performance of several conditional covariance estimators by examining certain characteristics of conditional correlation when the true correlation processes are observable. We simulate bivariate GARCH processes by considering two univariate Gaussian GARCH processes (e.g., Engle, 2002):

$$r_{1,t} = \sqrt{h_{1,t}}\varepsilon_{1,t},\ h_{1,t} = 0.01 + 0.05r_{1,t-1}^2 + 0.94h_{1,t-1},$$
$$r_{2,t} = \sqrt{h_{2,t}}\varepsilon_{2,t},\ h_{2,t} = 0.5 + 0.2r_{2,t-1}^2 + 0.5h_{2,t-1},$$

where

$$\left(\begin{array}{c} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{array}\right)\Bigg| \mathcal{F}_{t-1} \sim N\left( 0 \quad \left(\begin{array}{cc} 1 & \rho_t \\ \rho_t & 1 \end{array}\right)\right).$$

We consider four specifications of $\rho_t$ which are given in the following DGPs:

DGP1: $\rho_t = 0.5 + 0.4\cos(2\pi t/20)$.

DGP2: $\rho_t = 0.99 - 1.98/\left\{1 + \exp\left[0.5 \times \max\left(\varepsilon_{1,t-1}^2, \varepsilon_{2,t-1}^2\right)\right]\right\}$.

DGP3: $\rho_t = 0.3 \times 1\left(\varepsilon_{1,t-1}^2 \leq h_{1,t-1}\right) + 0.8 \times 1\left(\varepsilon_{1,t-1}^2 > h_{1,t-1}\right)$.

DGP4: $\rho_t = \rho_{2t}$ with probability 0.5 and $= \rho_{3t}$ with probability 0.5, where $\rho_{2t}$ and $\rho_{3t}$ are $\rho_t$ specified in DGP2 and DGP3, respectively.

DGP1 was also adopted by Engle (2002). DGPs 2-3 are motivated by the stylized fact in financial markets that conditional correlation in crisis periods is higher than that in tranquil periods. DGP3 also borrows idea of regime switching from Pelletier (2006). DGP 4 is the mixture of DGP 2 and DGP 3. For each DGP, we will simulate 1000 observations on $\mathbf{r}_t = (r_{1,t}.r_{2,t})'$. After throwing away the first 500 observations to avoid the starting-out effect, we are left with $T = 500$ observations, which represents roughly two-year daily data. The number of replications for each case is $M = 200$.

We will consider four parametric models for estimating the conditional correlation of $r_t$, namely the CCC, VC, SBEKK and DCC models reviewed in Section 2. In each case, we will also obtain our SCC estimators by choosing the conditioning variable $\mathbf{x}_t = r_{t-1}$. To obtain our SCC estimators, we need to choose both the kernel function and the bandwidth parameter. It is well known that the choice of kernel function $k(\cdot)$ is not important in nonparametric or semiparametric estimation. We will simply use the Gaussian kernel:

$$k(u) = \exp\left(-u^2/2\right)/\sqrt{2\pi}.$$

For the bandwidth, we follow the idea of grid-searching and set

$$h_j = c_j \widehat{\sigma}_j n^{-1/6},\ j = 1, 2,$$

where $\widehat{\sigma}_j$ is the sample standard deviation of $r_{j,t}$, and the optimal $c_j$ is chosen from 0.5, 1, 1.5,...,5 by minimizing loss function of the corresponding semiparametric model.

Similar to Engle (2002), for each DGP and each estimator we calculate the mean squared errors (MSE) across the $M$ replications

$$\text{MSE}(\widehat{\rho}_t) = \frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} \left(\widehat{\rho}_t^{(m)} - \rho_t^{(m)}\right)^2, \tag{4.1}$$

where $\rho_t^{(m)}$ and $\widehat{\rho}_t^{(m)}$ are the true conditional correlation and its estimates at time $t$ in the $m$th replication, respectively.

In Table 1, we compare the simulation results for the CCC, VC, SBEKK, DCC models and their semiparametric counterparts in terms of MSE($\widehat{\rho}_t$) and their improvement ratio, which is defined as

$$\text{ratio} = \left[\frac{\text{Loss}\,(\text{PCC}) - \text{Loss}\,(\text{SCC})}{\text{Loss}\,(\text{PCC})}\right] \times 100 \tag{4.2}$$

where Loss(SCC) and Loss(PCC) are the MSE[2] for the SCC estimator and for the start-up PCC estimator, respectively. We summarize some interesting findings below.

(1) Our SCC estimator always beat the PCC estimator that is used as its parametric start. Take DGP1 as an example. The DCC model by Engle (2002) has the second smallest MSE values (0.069078) among the parametric estimators; our semiparametric estimator can further decrease it to 0.068965.

(2) In terms of MSE, the rankings of semiparametric estimators are consistent with the rankings of parametric estimators. In DGP3, for instance, the performance of the parametric estimators in the CCC, VC, SBEKK and DCC models deteriorates in order, so does the performance of the semiparametric estimators with these corresponding models as the start-up models.

(3) There is no clear relationship between the relative ranking of parametric start-up models based on their MSE values and the improvement extent of the corresponding semiparametric estimators in terms of their improvement ratios. We did not observe that for the parametric start-up model being far away from the true model, its corresponding semiparametric estimator has more gain in terms of MSE compared with the case when the parametric start model is closer to the true model. In DGP 1, for example, the parametric model ranking with ascending MSE value is SBEKK, DCC, VC to CCC models, which is not the ascending improvement ratio ranking of DCC-NW, SBEKK-NW, VC-NW and CCC-NW models. Take DGP4 as another example. The range of improvement ratio varies from 3.417% to 5.116%. The improvement ratio for the SBEKK-NW model over the SBEKK model is the smallest whereas that for CCC-NW model over the CCC model is the largest.[3] This occurs whilst among the PCC models the SBEKK model has the largest MSE and the CCC model has the smallest MSE.

## 4.2 Empirical Analysis

Although many multivariate conditional covariance models are proposed, only a few are applied to high-dimensional case, for example, 7 stock indices in Ledoit, Santa-Clara, and Wolf (2003). To show our capacity of expanding to high dimension, we examine the data sets consisting of 30 stocks in the last two empirical applications in addition to the first example of bivariate stock indices. These three interesting financial daily time series are: the NASDAQ Composite Index and Standard & Poor's 500 Index (NASDAQ - SP) from January 2, 1990 to December 30, 1994 ($T = 1265$ observations); the first 30 stocks[4] existing since beginning (January 3, 1984) among the component stocks of Financial Times Stock Exchange 100 Index (FTSE) sorted by alphabet from January 4, 2000 to December

[2] Or VaR loss discussed in the next subsection.

[3] The SBEKK-NW model denotes the SCC model incorporating parametric SBEKK model and nonparametric Nadaraya-Watson model multiplicatively. Similar notation is used for the CCC-NW, VC-NW, and DCC-NW models.

[4] The stock name list is available upon request.

31, 2003 ($T = 1009$ observations); and the 30 stocks constituting the Dow Jones Industrial Average (DJIA) from January 4, 1993 to December 31, 1997 ($T = 1264$ observations). The first data sets is from Yahoo and the last two are from Datastream. For stationary and ease interpretation, we thus compute percentage returns as log returns multiplied by 100. We consider the entire sample period for IS estimators and the corresponding standardized residuals are bootstrapped to compute the $p$-value for Value-at-Risk (VaR) calculation later. We split the samples at day $R$, the last day of the second last year and apply the "fixed scheme" to do one-day-ahead conditional covariance matrix forecast throughout a whole year beyond day $R$: estimating parameters based on information set $F_R$ and fixing these estimated parameter values to make forecasts throughout the forecasting period with $P = T - R - 1 = 252$, 251 and 253 observations in three cases, respectively. Thus, for example, the IS period for DJIA goes from January 5, 1993 to December 31, 1997 and the OoS period goes from January 2, 1997 to December 31, 1997.

Any discussions on conditional mean and its relationship with conditional covariance are outside the scope of the paper. For these data sets, we assume the conditional means are zero based on efficient market hypothesis and sample-mean filter is used. We choose the kernel function and the bandwidth as we do in the simulation. Although there is no consensus among the Finance profession on the identity of the common factors, - how many and which ones are?- stock index is always the first factor being picked up. Thus rather than taking the lagged studied variables as in the simulation and in the first empirical application, we select the one-day lag percentage return of DJIA Index (FTSE 100 Index) as the state variable for the 30-stock set in the second (third) application.

We consider two types of criterion functions to judge the relative fitting and predictive ability of various conditional covariance models in terms of portfolios' certain characteristics. One is the modified mean squared error that is adapted to our framework because the true conditional covariance matrix is not observable. Zangari (1997) addresses the advantage of focusing on the volatility $h_t^y$ of the aggregate portfolio $y_t \equiv \boldsymbol{\omega}'\mathbf{r}_t$ instead of the conditional covariance matrix $\mathbf{H}_t$, where $h_t^y = \boldsymbol{\omega}'\mathbf{H}_t\boldsymbol{\omega}$ and $\boldsymbol{\omega}$ is a weight vector. When comparing the predictability of univariate GARCH models, Awartani and Corradi (2005) substitute the unobservable volatility by the squared observed returns because of the rank-preserving property of this substitution under the MSE loss function. They conclude that both this squared returns and realized volatility are good proxies of the unobservable volatility when interested in model comparisons. Because of the unavailability of intraday returns, the proxy in the criteria to compare volatility and correlation forecasts in Pelletier (2006) is the cross-product of daily return instead of cumulative cross-product of intraday returns over the forecast horizon. Following them, we compare various models by calculating the fitting and predictive measures, $\mathrm{MSE}_{\mathrm{IS}}^j$ and

$\mathrm{MSE}_{\mathrm{OoS}}^{j}$[5], for model $j$, as

$$\mathrm{MSE}_{\mathrm{OoS}}^{j} = \frac{1}{P}\sum_{t=R}^{T-2}\left(\boldsymbol{\omega}_{t+1}'\widehat{\mathbf{H}}_{t+1}^{j}\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_{t+1}'\mathbf{r}_{t+1}\mathbf{r}_{t+1}'\boldsymbol{\omega}_{t+1}\right)^{2}, \tag{4.3}$$

where $\widehat{\mathbf{H}}_{t+1}^{j}$ is the one-step-ahead forecaster of $\mathbf{H}_{t+1}$ at time $t$ from model $j$.

The second loss is based on the portfolios' VaR. The Basel Committee on Banking Supervision uses VaR to estimate the risk exposure of financial institutes for a ten-day holding period and 99% coverage ($\alpha = 1\%$). Denote the VaR forecast of the weighted portfolio with tail probability $\alpha$ from model $j$ within our framework as

$$\mathrm{VaR}_{\mathrm{OoS},t+1}^{\alpha,j} = \Phi_{\alpha}^{j}\sqrt{\boldsymbol{\omega}_{t+1}'\widetilde{\mathbf{H}}_{t+1}^{j}\boldsymbol{\omega}_{t+1}}, \tag{4.4}$$

where $\Phi_{\alpha}^{j}$ is the quantile of cumulative distribution function of weighted portfolio at tail probability $\alpha \in (0,1)$ from model $j$. Apart from adopting the quantiles of standard normal distribution, Bauwens and Laurent (2005) use a Monte Carlo simulation and HDF employ the quantiles of the standardized IS portfolio returns. Instead, we compute $\Phi_{\alpha}^{j}$ by repeated bootstrap sampling the weighted standardized residuals over the entire samples. Hall (1986) explains theoretically few bootstrap sampling replications could produce satisfactory results, thus we compute bootstrap critical values via 100 replications of the bootstrap sampling process. Note that the weighted sum of the underlying asset returns' VaR is not equal to the weighted portfolio's VaR. The OoS VaR loss function for model $j$, which is the check function of Koenker and Bassett (1978), is

$$Q_{\mathrm{OoS}}^{j} = \frac{1}{P}\sum_{t=R}^{T-2}(\alpha - \mathbf{1}(y_{t+1} < \mathrm{VaR}_{\mathrm{OoS},t+1}^{\alpha,j}))(y_{t+1} - \mathrm{VaR}_{\mathrm{OoS},t+1}^{\alpha,j}), \tag{4.5}$$

where $\mathbf{1}(\cdot)$ is an indicator function. The tail characteristics at $\alpha = 1\%$ and 5% of the portfolio constructed by the equal weight and the minimum variance weight are examined. For the equally weighted portfolio, the constant equal weight is $\boldsymbol{\omega}_t = k^{-1}\mathbf{e}$, where $\mathbf{e}$ is a $k \times 1$ vector of ones; and for the minimum variance portfolio, the weight is $\boldsymbol{\omega}_t = \mathbf{H}_t^{-1}\mathbf{e}/\left(\mathbf{e}'\mathbf{H}_t^{-1}\mathbf{e}\right)$.

The IS and OoS performance measures of different conditional covariance models over these empirical datasets are presented in Table 2-4, with the equal weight portfolio results in Panel A and the minimum variance portfolio results in Panel B. For each pair of parametric start-up model and the corresponding semiparametric model, the improvement ratio defined as loss difference over parametric model's loss is reported in percentage for both IS and OoS cases.

Some important findings are presented as: (1) For both criterion functions, our semiparametric model can always reduce the loss values of the start-up parametric model no matter which weight

[5] To save space, we only present formulae for $\mathrm{MSE}_{\mathrm{OoS}}^{j}$, $\mathrm{VaR}_{\mathrm{OoS},t+1}^{\alpha,j}$ and $Q_{\mathrm{OoS}}^{\alpha,j}$. It is not difficult to derive the formulae for $\mathrm{MSE}_{\mathrm{IS}}^{j}$, $\mathrm{VaR}_{\mathrm{IS},t}^{\alpha,j}$ and $Q_{\mathrm{IS}}^{\alpha,j}$.

or which sample period (IS or OoS) we are interested in. The improvement is significant across all these datasets. For the NASDAQ-SP dataset, the range of improvement ratio ranges from 0.586% (OoS VaR loss at 1% for CCC vs. CCC-NW models with minimum variance weight) to 48.373% (IS MSE for SBEKK vs. SBEKK-NW model with minimum variance weight); for the FTSE data set, the range is from 0.460% (IS MSE for DCC vs. DCC-NW models with equal weight) to 81.446% (IS MSE for SBEKK vs. SBEKK-NW models with minimum variance weight); and for the DJIA data set, the range is from 2.776% (OoS VaR loss at 5% for DCC vs. DCC-NW models with equal weight) to 80.317% (IS MSE for SBEKK vs. SBEKK-NW models with minimum variance weight).

(2) There exists no semiparametric model that is universally the best across different datasets, weighting methods or loss functions. While the SBEKK-NW model has the smallest OoS values across the loss functions for the minimum variance DJIA portfolio, its OoS MSE is bigger than that of the VC-NW model for the equal weight DJIA portfolio. For the minimum variance DJIA portfolio, the DCC-NW model beats all other models in terms of IS losses, but we do not observe this dominance for the equal weight DJIA portfolio.

(3) For the same conditional covariance model, the minimum variance portfolio always outperforms the equal weight portfolio in terms of the loss functions we examine across all datasets. For the VC model, for example, the VaR loss at 1% of equal weighted DJIA portfolio is 0.144, bigger than that of minimum variance DJIA portfolio, 0.085; and for the DCC-NW model, the VaR loss at 5% of the minimum variance NASDAQ-SP portfolio is 0.072, smaller than that of the equal weight NASDAQ-SP portfolio, 0.087.

(4) The improvement ratio by semiparametric methodology does in favour of the more extreme case for high-dimensional portfolio no matter which portfolio and which start-up parametric model we choose: the improvement ratios of VaR loss at tail probability of 1% are bigger than those at 5%. To the minimum variance DJIA portfolio, the DCC-NW model's improvement ratio for IS VaR loss when $\alpha = 1\%$ is 70.767%, 23.690% higher than the corresponding value when $\alpha = 5\%$, 47.077%; and for the equal weight DJIA portfolio, this difference is 17.859%. We observe the same phenomena for FTSE portfolio: the improvement ratio difference of VC-NW between OoS VaR losses at 1% and at 5% is 18.186% for minimum variance weight and 22.127% for equal weight.

# 5 CONCLUSION

In this paper, we propose a new semiparametric kernel-based modeling methodology for conditional covariance matrix, which applies Nadaraya-Watson type estimator to extract the information hidden in the standardized residuals from the parametric MGARCH models. Our SCC model combines parametric model and nonparametric model in a multiplicative way. For every parametric MGARCH model, there is a semiparametric counterpart, which is robust to two common misspecifications of the conventional parametric MGARCH models: multivariate normal distribution and linear functional form. We show that our semiparametric estimator is consistent and asymptotically normal under some regularity conditions. To examine the finite sample performance of our semiparametric estimators, we conduct a small set of Monte Carlo experiments inspired by the asymmetric correlations of financial time series data. We find that the nonparametric correction at the second stage of our SCC estimator can indeed improve the finite sample performance of the parametric MGARCH estimators in terms of MSE. Empirical applications on stock indices, and high-dimensional stock portfolio are quite encouraging: our new SCC models outperform their ancestors, the parametric MGARCH models, including the DCC model of Engle (2002), in reducing the values of MSE and VaR loss for both IS fitting and OoS forecasting conditional covariance matrix.

# ACKNOWLEDGEMENTS

The authors gratefully thank Serena Ng, the associate editor, and an anonymous referee for their many insightful comments that have led great improvement of the presentation. All errors are the authors' responsibilities. The second author gratefully acknowledges the financial support from the NSFC under the grant numbers 70501001 and 70601001. The third author acknowledges the financial support from the academic senate, UCR.

# Appendix

## A Proof of the Main Results

We use $C$ to signify a generic constant whose exact value may vary from case to case, and $a'$ to denote the transpose of $a$. Let $\widehat{f}(\mathbf{x}) = T^{-1}\sum_{s=1}^{T} K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})$, and

$$\widetilde{\mathbf{H}}_{np}(\mathbf{x}) = T^{-1}\sum_{s=1}^{T} \mathbf{e}_s\mathbf{e}_s' K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})/\widehat{f}(\mathbf{x}).$$

The following two lemmas are needed for the proof of Theorem 3.1.

**Lemma A.1** *Under Assumptions A1-A7,*

$$\sqrt{T\mathbf{h}!}\left\{vech\left(\widetilde{\mathbf{H}}_{np}(\mathbf{x})\right) - vech(\mathbf{H}_{np}(\mathbf{x})) - vech(\mathbf{B}(x))\right\} \xrightarrow{d} N\left(0, \mu_{02}^{q} f(\mathbf{x})^{-1} D_k^{+}\mathbf{\Omega}(\mathbf{x}) D_k^{+\prime}\right)$$

*where recall* $\mathbf{h}!$=$h_1 ... h_q$, *and* $\mathbf{\Omega}(\mathbf{x})$ *and* $\mathbf{B}(x)$ *are defined in Theorem 3.1.*

**Proof.** Let $W_{Tijs} = K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})e_{is}e_{js}$ and $W_{Tij} = T^{-1}\sum_{s=1}^{T} W_{Tijs}$, where $e_{is}$ is the $i$th element of $\mathbf{e}_s$. Define two $k(k+1)/2-$vectors:

$$\begin{aligned} W_{Ts} &= (W_{T11s}, W_{T21s} ..., W_{Tk1s}, W_{T22s}, ..., W_{Tk2s}, ..., W_{Tkks})' \\ W_T &= (W_{T11}, W_{T21} ..., W_{Tk1}, W_{T22}, ..., W_{Tk2}, ..., W_{Tkk})'. \end{aligned}$$

Clearly, $W_T = T^{-1}\sum_{s=1}^{T} W_{Ts}$. The statistic $W_{Tij}/\widehat{f}(\mathbf{x})$ estimates the $(i,j)$th element of $H_{np}(\mathbf{x})$ by using the "data" $\{\mathbf{e}_t, \mathbf{x}_t\}$. Let $Z_{Ts} = (\mathbf{h}!/T)^{1/2}(W_{Ts} - E(W_{Ts}))$ and $Z_T = \sum_{s=1}^{T} Z_{Ts}$. Write

$$\begin{aligned} W_T &= T^{-1}\sum_{s=1}^{T} E(W_{Ts}) + T^{-1}\sum_{s=1}^{T}(W_{Ts} - E(W_{Ts})) \\ &= T^{-1}\sum_{s=1}^{T} E(W_{Ts}) + (T\mathbf{h}!)^{-1/2}\sum_{s=1}^{T} Z_{Ts} \end{aligned}$$

The first term contributes to the bias of $\widetilde{\mathbf{H}}_{np}(\mathbf{x})$ whereas the second term contributes to the variance of $\widetilde{\mathbf{H}}_{np}(\mathbf{x})$. The proof will be completed by proving the following claims:

$$\widehat{f}(\mathbf{x}) \xrightarrow{p} f(\mathbf{x}), \tag{A.1}$$

$$T^{-1}\sum_{s=1}^{T} E(W_{Ts}) = f(\mathbf{x})\,\mathrm{vech}(\mathbf{H}_{np}(\mathbf{x})) + f(\mathbf{x})\,\mathrm{vech}(\mathbf{B}(\mathbf{x})) + o_P(\|\mathbf{h}\|^2). \tag{A.2}$$

and

$$Z_T = \sum_{s=1}^{T} Z_{Ts} \xrightarrow{d} N\left(0, \mu_{02}^{q} f(\mathbf{x}) D_k^{+}\mathbf{\Omega}(\mathbf{x}) D_k^{+\prime}\right). \tag{A.3}$$

(A.1) follows from standard results in kernel density estimation. Using standard arguments for analyzing the bias of the Nadaraya-Watson estimator, we have

$$E\left(W_{Tijs}\right) = E\left[K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})e_{it}e_{jt}\right] = f\left(\mathbf{x}\right)\left[\mathbf{H}_{np,ij}\left(\mathbf{x}\right) + \mathbf{B}_{ij}\left(\mathbf{x}\right)\right] + o_P(\|\mathbf{h}\|^2)$$

where

$$\mathbf{B}_{ij}\left(\mathbf{x}\right) = \frac{\mu_{21}}{2f\left(\mathbf{x}\right)}\sum_{l=1}^{q}\left[2\frac{\partial f\left(\mathbf{x}\right)}{\partial x_l}\frac{\partial \mathbf{H}_{np,ij}\left(\mathbf{x}\right)}{\partial x_l} + f\left(\mathbf{x}\right)\frac{\partial^2 \mathbf{H}_{np,ij}\left(\mathbf{x}\right)}{\partial x_l \partial x_l}\right]h_l^2.$$

Thus (A.2) follows by the stationarity assumption. To show (A.3), let $\mathbf{c} = (c_{11}, c_{21}..., c_{k1}, c_{22}, ..., c_{k2}, ..., c_{kk})'$ denote a $k\left(k+1\right)/2-$vector of bounded constants such that $\|\mathbf{c}\| = 1$. By the Cramér-Wold device, it suffices to show

$$\mathbf{c}' Z_T = \sum_{s=1}^{T}\mathbf{c}' Z_{Ts} \xrightarrow{d} N\left(0, \mu_{02}^q f\left(\mathbf{x}\right)\mathbf{c}' D_k^+ \mathbf{\Omega}\left(\mathbf{x}\right) D_k^{+\prime}\mathbf{c}\right). \quad \text{(A.4)}$$

By construction, $E\left(Z_T\right) = 0$, and

$$\mathrm{Var}\left(\mathbf{c}' Z_T\right) = T^{-1}\mathbf{h}!\sum_{t=1}^{T}\mathrm{Var}\left(\mathbf{c}' W_{Tt}\right) + 2T^{-1}\mathbf{h}!\sum\sum_{1\le s<t\le T}\mathrm{Cov}\left(\mathbf{c}' W_{Ts}, \mathbf{c}' W_{Tt}\right) \equiv A_1 + A_2. \quad \text{(A.5)}$$

We calculate $A_1$ and $A_2$ in turn.

$$\begin{aligned} A_1 &= T^{-1}\mathbf{h}!\sum_{t=1}^{T}\mathrm{Var}\left(\mathbf{c}' W_{Tt}\right) \\ &= \sum\sum_{1\le j\le i\le k}\sum\sum_{1\le m\le l\le k} c_{ij}c_{lm}\left[T^{-1}\mathbf{h}!\sum_{t=1}^{T}E\left[K_{\mathbf{h}}^2(\mathbf{x}_t - \mathbf{x})\mathrm{Cov}\left(\varrho_{ij,t}, \varrho_{lm,t}|\mathbf{x}_t = \mathbf{x}\right)\right]\right] \\ &= \mu_{02}^q f\left(\mathbf{x}\right)\sum\sum_{1\le j\le i\le k}\sum\sum_{1\le m\le l\le k} c_{ij}c_{lm}\omega_{ij,lm}\left(\mathbf{x}\right) + O\left(\|\mathbf{h}\|\right) \\ &= \mu_{02}^q f\left(\mathbf{x}\right)\mathbf{c}' D_k^+\mathbf{\Omega}\left(\mathbf{x}\right)D_k^{+\prime}\mathbf{c} + O\left(\|\mathbf{h}\|\right), \end{aligned} \quad \text{(A.6)}$$

where $\varrho_{ij,t} = e_{it}e_{jt}$ and $\omega_{ij,lm}\left(\mathbf{x}\right) = \mathrm{Cov}\left(\varrho_{ij,t}, \varrho_{lm,t}|\mathbf{x}_t = \mathbf{x}\right)$. To calculate $A_2$, write

$$\begin{aligned} A_2 &= 2T^{-1}\mathbf{h}!\sum\sum_{1\le s<t\le T}\sum\sum_{1\le j\le i\le k}\sum\sum_{1\le m\le l\le k} c_{ij}c_{lm}\mathrm{Cov}\left(W_{Tijs}, W_{Tlmt}\right) \\ &= 2\mathbf{h}!\sum_{t=2}^{T}\left(1 - \frac{j}{T}\right)\sum\sum_{1\le j\le i\le k}\sum\sum_{1\le m\le l\le k} c_{ij}c_{lm}\mathrm{Cov}\left(W_{Tij1}, W_{Tlmt}\right). \end{aligned} \quad \text{(A.7)}$$

Noting that even though $\{\mathbf{v}_t\}$ is a m.d.s., this does not ensure that $\mathrm{Cov}(W_{Tijs}, W_{Tlmt}) = 0$ for $s \ne t$. To bound the right hand side of (A.7), we split it into two terms as follows

$$\sum_{t=2}^{T}\left|\mathrm{Cov}\left(W_{Tij1}, W_{Tlmt}\right)\right| = \sum_{t=2}^{d_T}\left|\mathrm{Cov}\left(W_{Tij1}, W_{Tlmt}\right)\right| + \sum_{t=d_T+1}^{T}\left|\mathrm{Cov}\left(W_{Tij1}, W_{Tlmt}\right)\right| \equiv J_1 + J_2, \quad \text{(A.8)}$$

where $d_T$ is a sequence of positive integers such that $d_T\mathbf{h}! \to 0$ as $T \to \infty$. Since for any $t > 1, \left|E\left(W_{Tij1}W_{Tlmt}\right)\right| = O\left(1\right)$,

$$J_1 = O\left(d_n\right). \tag{A.9}$$

For $J_2$, by the Davydov's inequality (e.g., Hall and Heyde, 1980, p. 278; Bosq, 1996, p.19), we have

$$\begin{aligned} \left|\mathrm{Cov}\left(W_{Tij1}W_{Tlmt}\right)\right| &\leq C\left[\alpha\left(t-1\right)\right]^{\delta/(2+\delta)} \sup_{i,j}\left(E\left|W_{Tij1}\right|^{2+\delta}\right)^{2/(2+\delta)} \\ &\leq C\left(\mathbf{h}!\right)^{-(2+2\delta)/(2+\delta)}\left[\alpha\left(t-1\right)\right]^{\delta/(2+\delta)}. \end{aligned}$$

So by Assumption A.1,

$$\begin{aligned} J_2 &\leq C\left(\mathbf{h}!\right)^{-(2+2\delta)/(2+\delta)} \sum_{t=d_T+1}^{T}\left[\alpha\left(t-1\right)\right]^{\delta/(2+\delta)} \\ &\leq C\left(\mathbf{h}!\right)^{-(2+2\delta)/(2+\delta)} d_T^{-a}\sum_{t=d_T}^{\infty} t^a\left[\alpha\left(t\right)\right]^{\delta/(2+\delta)} = o\left(\left(\mathbf{h}!\right)^{-1}\right), \end{aligned} \tag{A.10}$$

by choosing $d_T$ such that $d_T^a\left(\mathbf{h}!\right)^{\delta/(2+\delta)} \to \infty$. The last condition can be simultaneously met with $d_T\mathbf{h}! \to 0$ for a well chosen sequence $\{d_T\}$ because $a > \delta/\left(2+\delta\right)$ by Assumptions A.1 and A.7. (A.7)-(A.10) imply that

$$A_2 = O\left(d_n\mathbf{h}!\right) + o\left(1\right) = o\left(1\right).$$

Hence,

$$\mathrm{Var}\left(\mathbf{c}'Z_T\right) = \mu_{02}^q f\left(\mathbf{x}\right)\mathbf{c}'D_k^+\mathbf{\Omega}\left(\mathbf{x}\right)D_k^{+\prime}\mathbf{c} + o\left(1\right).$$

Using the standard Doob's small-block and large-block technique, we can finish the rest of the normality proof of (A.4) by following the arguments of Cai, Fan and Yao (2000, pp. 954-955) or Cai and Ould-Saïd (2003, pp. 446-448). ■

**Lemma A.2** *Under Assumptions A1-A7,*

$$vech\left(\widehat{\mathbf{H}}_{np}\left(\mathbf{x}\right)\right) - vech\left(\widetilde{\mathbf{H}}_{np}\left(\mathbf{x}\right)\right) = o_P\left(\left(T\mathbf{h}!\right)^{-1/2}\right).$$

**Proof.** Let $\Delta\left(\mathbf{x}\right) = [\mathrm{vec}(\widehat{\mathbf{H}}_{np}\left(\mathbf{x}\right)) - \mathrm{vec}(\widetilde{\mathbf{H}}_{np}\left(\mathbf{x}\right))]\widehat{f}\left(\mathbf{x}\right)$. Noting that $\widehat{f}\left(\mathbf{x}\right) \xrightarrow{p} f\left(\mathbf{x}\right) > 0$ and $\mathrm{vech}(\mathbf{A}) = D_k^+\mathrm{vec}(\mathbf{A})$ for any symmetric $k\times k$ matrix $\mathbf{A}$, it suffices to show that $\Delta\left(\mathbf{x}\right) = o_P((T\mathbf{h}!)^{-1/2})$. By the first order expansion,

$$\widehat{\mathbf{e}}_t = \mathbf{e}_t(\widehat{\theta}) = \mathbf{H}_{p,t}^{-1/2}\left(\widehat{\theta}\right)\mathbf{r}_t = \mathbf{e}_t + \boldsymbol{\xi}_t\left(\overline{\theta}\right)\left(\widehat{\theta} - \theta_*\right) \tag{A.11}$$

where recall $\boldsymbol{\xi}_t(\theta) = \partial \mathbf{e}_t(\theta)/\partial\theta'$, and $\overline{\theta}$ lies between $\widehat{\theta}$ and $\theta_*$. By Assumptions A2-A3, $\overline{\theta} \xrightarrow{p} \theta_*$. So

$$\begin{aligned}
\Delta(\mathbf{x}) &= \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\mathrm{vec}\left[\widehat{\mathbf{e}}_t\widehat{\mathbf{e}}_t' - \mathbf{e}_t\mathbf{e}_t'\right] \\
&= \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\mathrm{vec}\left[\boldsymbol{\xi}_t\left(\overline{\theta}\right)\left(\widehat{\theta}-\theta_*\right)\left(\widehat{\theta}-\theta_*\right)'\xi_t\left(\overline{\theta}\right)'\right] \\
&\quad + \frac{2}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\mathrm{vec}\left[\mathbf{e}_t\left(\widehat{\theta}-\theta_*\right)'\xi_t\left(\overline{\theta}\right)'\right] \\
&= \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\left(\boldsymbol{\xi}_t\left(\overline{\theta}\right)\otimes\boldsymbol{\xi}_t\left(\overline{\theta}\right)\right)\mathrm{vec}\left[\left(\widehat{\theta}-\theta_*\right)\left(\widehat{\theta}-\theta_*\right)'\right] \\
&\quad + \frac{2}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\left(\boldsymbol{\xi}_t\left(\overline{\theta}\right)\otimes\mathbf{e}_t\right)\left(\widehat{\theta}-\theta_*\right) \\
&\equiv \Delta_1(\mathbf{x}) + 2\Delta_2(\mathbf{x}).
\end{aligned}$$

By the triangle inequality, Markov inequality, and Assumptions A4-A7,

$$\begin{aligned}
\|\Delta_1(\mathbf{x})\| &\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\left\|\left(\boldsymbol{\xi}_t\left(\overline{\theta}\right)\otimes\boldsymbol{\xi}_t\left(\overline{\theta}\right)\right)\mathrm{vec}\left[\left(\widehat{\theta}-\theta_*\right)\left(\widehat{\theta}-\theta_*\right)'\right]\right\| \\
&\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\left\|\boldsymbol{\xi}_t\left(\overline{\theta}\right)\right\|^2\left\|\widehat{\theta}-\theta_*\right\|^2 \\
&\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\overline{D}_t^2\left\|\widehat{\theta}-\theta_*\right\|^2 = O_P\left(\frac{1}{T\mathbf{h}!}\right),
\end{aligned}$$

and

$$\begin{aligned}
\|\Delta_2(\mathbf{x})\| &\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\left\|\left(\boldsymbol{\xi}_t\left(\overline{\theta}\right)\otimes\mathbf{e}_t\right)\left(\widehat{\theta}-\theta_*\right)\right\| \\
&\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\left\|\boldsymbol{\xi}_t\left(\overline{\theta}\right)\right\|\|\mathbf{e}_t\|\left\|\widehat{\theta}-\theta_*\right\| \\
&\le \frac{1}{T}\sum_{t=1}^{T} K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\overline{D}_t\|\mathbf{e}_t\|\left\|\widehat{\theta}-\theta_*\right\| = O_P\left(T^{-1/2}\right).
\end{aligned}$$

Consequently, $\Delta(\mathbf{x}) = O_P((T\mathbf{h}!)^{-1} + T^{-1/2}) = o_P((T\mathbf{h}!)^{-1/2})$.■

**Proof of Theorem 3.1**

The result follows from Lemmas A.1-A.2.■

**Proof of Corollary 3.2**

By Assumptions A3-A5, $\widehat{\mathbf{H}}_{p,t} = \mathbf{H}_{p,t}^{1/2}(\widehat{\theta}) = \mathbf{H}_{p,t}^{1/2} + o_P(1)$. By Theorem 3.1, $\widehat{\mathbf{H}}_{np,t} = \widehat{\mathbf{H}}_{np}(\mathbf{x}_t) = \mathbf{H}_{np,t} + o_P(1)$. It follows from the Slutsky theorem that

$$\widehat{\mathbf{H}}_{sp,t} = \widehat{\mathbf{H}}_{p,t}^{1/2}\widehat{\mathbf{H}}_{np,t}\widehat{\mathbf{H}}_{p,t}^{1/2} = \mathbf{H}_{p,t}^{1/2}\mathbf{H}_{np,t}\mathbf{H}_{p,t}^{1/2} + o_P(1) = \mathbf{H}_t + o_P(1),$$

and

$$\widehat{\mathbf{H}}_{sp,t}^{*} = \mathbf{H}_t^{*} + o_P(1),$$

where $\mathbf{H}_t^{*}$ is a diagonal matrix with the square roots of the diagonal elements of $\mathbf{H}_t$ on its diagonal. Hence

$$\widehat{\mathbf{R}}_{sp,t} = \left(\widehat{\mathbf{H}}_{sp,t}^{*}\right)^{-1}\widehat{\mathbf{H}}_{sp,t}\left(\widehat{\mathbf{H}}_{sp,t}^{*}\right)^{-1} \xrightarrow{p} \left(\mathbf{H}_t^{*}\right)^{-1}\mathbf{H}_t\left(\mathbf{H}_t^{*}\right)^{-1} = \mathbf{R}_t.$$

To show (ii), noting that by Assumptions A3-A5,

$$\begin{aligned}
\widehat{\mathbf{H}}_{sp,t} - \mathbf{H}_t &= \widehat{\mathbf{H}}_{p,t}^{1/2}\widehat{\mathbf{H}}_{np,t}\widehat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\mathbf{H}_{np}(\mathbf{x}_t)\mathbf{H}_{p,t}^{1/2} \\
&= \mathbf{H}_{p,t}^{1/2}\left(\widehat{\mathbf{H}}_{np}(\mathbf{x}_t) - \mathbf{H}_{np}(\mathbf{x}_t)\right)\mathbf{H}_{p,t}^{1/2} + \left\{\left(\widehat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\right)\widehat{\mathbf{H}}_{np,t}\left(\widehat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\right)\right. \\
&\quad \left. + \left(\widehat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\right)\widehat{\mathbf{H}}_{np,t}\mathbf{H}_{p,t}^{1/2} + \mathbf{H}_{p,t}^{1/2}\widehat{\mathbf{H}}_{np,t}\left(\widehat{\mathbf{H}}_{p,t}^{1/2} - \mathbf{H}_{p,t}^{1/2}\right)\right\} \\
&= \mathbf{H}_{p,t}^{1/2}\left(\widehat{\mathbf{H}}_{np}(\mathbf{x}_t) - \mathbf{H}_{np}(\mathbf{x}_t)\right)\mathbf{H}_{p,t}^{1/2} + O_p\left(T^{-1/2}\right),
\end{aligned}$$

we have

$$\begin{aligned}
&\sqrt{T\mathbf{h}!}\left[\mathrm{vech}\left(\widehat{\mathbf{H}}_{sp,t}\right) - \mathrm{vech}(\mathbf{H}_t)\right] \\
&= \sqrt{T\mathbf{h}!}D_k^{+}\left[\mathrm{vec}\left(\widehat{\mathbf{H}}_{sp,t}\right) - \mathrm{vec}(\mathbf{H}_t)\right] \\
&= \sqrt{T\mathbf{h}!}D_k^{+}\mathrm{vec}\left(\mathbf{H}_{p,t}^{1/2}\left(\widehat{\mathbf{H}}_{np}(\mathbf{x}_t) - \mathbf{H}_{np}(\mathbf{x}_t)\right)\mathbf{H}_{p,t}^{1/2}\right) + o_P(1) \\
&= \sqrt{T\mathbf{h}!}D_k^{+}\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)\mathrm{vec}\left(\widehat{\mathbf{H}}_{np}(\mathbf{x}_t) - \mathbf{H}_{np}(\mathbf{x}_t)\right) + o_P(1) \\
&= \sqrt{T\mathbf{h}!}D_k^{+}\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)D_k\mathrm{vech}\left(\widehat{\mathbf{H}}_{np}(\mathbf{x}_t) - \mathbf{H}_{np}(\mathbf{x}_t)\right) + o_P(1).
\end{aligned}$$

Then by Theorem 3.1,

$$\sqrt{T\mathbf{h}!}\left[\mathrm{vech}\left(\widehat{\mathbf{H}}_{sp,t}\right) - \mathrm{vech}(\mathbf{H}_t) - \overline{\mathbf{B}}_t(\mathbf{x}_t)\right] \xrightarrow{d} MN\left(0, \mu_{02}^{q} f(\mathbf{x}_t)^{-1}\overline{\overline{\mathbf{\Omega}}}_t(\mathbf{x})\right)$$

where

$$\begin{aligned}
\overline{\mathbf{B}}_t(\mathbf{x}) &= D_k^{+}\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)D_k\mathrm{vech}(\mathbf{B}(\mathbf{x})) = D_k^{+}\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)\mathrm{vec}(\mathbf{B}(\mathbf{x})) \\
&= D_k^{+}\mathrm{vec}\left(\mathbf{H}_{p,t}^{1/2}\mathbf{B}(\mathbf{x})\mathbf{H}_{p,t}^{1/2}\right) = \mathrm{vech}\left(\mathbf{H}_{p,t}^{1/2}\mathbf{B}(\mathbf{x})\mathbf{H}_{p,t}^{1/2}\right)
\end{aligned}$$

by the definitions of vech, vec, $D_k$, and $D_k^{+}$ and the fact that $(\mathbf{A}\otimes\mathbf{A})\mathrm{vec}(\mathbf{B}(\mathbf{x})) = \mathrm{vec}(\mathbf{A}\mathbf{B}(\mathbf{x})\mathbf{A})$ for any $k\times k$ matrix $\mathbf{A}$, and

$$\begin{aligned}
\overline{\overline{\mathbf{\Omega}}}_t(\mathbf{x}) &= D_k^{+}\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)D_kD_k^{+}\mathbf{\Omega}(\mathbf{x})D_k^{+\prime}D_k^{\prime}\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)\left(D_k^{+}\right)^{\prime} \\
&= D_k^{+}N_k\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)\mathbf{\Omega}(\mathbf{x})\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)N_k\left(D_k^{+}\right)^{\prime} \\
&= D_k^{+}\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)\mathbf{\Omega}(\mathbf{x})\left(\mathbf{H}_{p,t}^{1/2}\otimes\mathbf{H}_{p,t}^{1/2}\right)\left(D_k^{+}\right)^{\prime} = D_k^{+}\overline{\mathbf{\Omega}}_t(\mathbf{x})\left(D_k^{+}\right)^{\prime}
\end{aligned}$$

by the fact that $N_k \equiv D_kD_k^{+}$ is symmetric, $N_kD_k = D_k$, $N_kD_k^{+\prime} = D_k^{+\prime}$, and $N_k(\mathbf{A}\otimes\mathbf{A}) = (\mathbf{A}\otimes\mathbf{A})N_k$ for any $k\times k$ matrix $\mathbf{A}$. ■

**Proof of Theorem 3.3**

Let $\boldsymbol{\iota}_i$ denote a $k \times 1$ vector that has 1 in the $i$th row and 0 elsewhere. Then

$$\widehat{e}_{it} = \boldsymbol{\iota}_i' \widehat{\mathbf{e}}_t = \boldsymbol{\iota}_i' \widehat{\mathbf{H}}_{p,t}^{-1/2} \mathbf{r}_t = \boldsymbol{\iota}_i' \mathbf{H}_{p,t}^{-1/2} \mathbf{r}_t + \boldsymbol{\iota}_i' \left( \widehat{\mathbf{H}}_{p,t}^{-1/2} - \mathbf{H}_{p,t}^{-1/2} \right) \mathbf{r}_t = e_{it} + \nu_{it},$$

where $\nu_{it} = \boldsymbol{\iota}_i' \left( \widehat{\mathbf{H}}_{p,t}^{-1/2} - \mathbf{H}_{p,t}^{-1/2} \right) \mathbf{r}_t$. Note that for notational simplicity we have suppressed the dependence of $\nu_{it}$ on the sample size $T$. It follows that

$$\frac{1}{T} \sum_{t=1}^{T} \left( \widehat{e}_{it} \widehat{e}_{jt} - \delta_{ij} \right) K_{\mathbf{h}} \left( \mathbf{x}_t - \mathbf{x} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ \left( e_{it} + \nu_{jt} \right) \left( e_{it} + \nu_{jt} \right) - \delta_{ij} \right] K_{\mathbf{h}} \left( \mathbf{x}_t - \mathbf{x} \right) = \sum_{l=1}^{4} A_{ij,l} \left( \mathbf{x} \right),$$

and

$$\frac{1}{T^2} \sum_{t=1}^{T} \left( \widehat{e}_{it} \widehat{e}_{jt} - \delta_{ij} \right)^2 \overline{K}_{\mathbf{h}} \left( \mathbf{0} \right) = \frac{1}{T} \sum_{t=1}^{T} \left[ \left( e_{it} + \nu_{it} \right) \left( e_{jt} + \nu_{jt} \right) - \delta_{ij} \right]^2 \overline{K}_{\mathbf{h}} \left( \mathbf{0} \right) = \sum_{l=1}^{4} B_{ij,l},$$

where

$$\begin{array}{ll}
A_{ij,1} \left( \mathbf{x} \right) = \frac{1}{T} \sum_{t=1}^{T} \left( e_{it} e_{jt} - \delta_{ij} \right) K_{\mathbf{h}} \left( \mathbf{x}_t - \mathbf{x} \right), & B_{ij,1} = \frac{1}{T} \sum_{t=1}^{T} \left( e_{it} e_{jt} - \delta_{ij} \right)^2 \overline{K}_{\mathbf{h}} \left( \mathbf{0} \right), \\
A_{ij,2} \left( \mathbf{x} \right) = \frac{1}{T} \sum_{t=1}^{T} \nu_{it} \nu_{jt} K_{\mathbf{h}} \left( \mathbf{x}_t - \mathbf{x} \right), & B_{ij,2} = \frac{1}{T} \sum_{t=1}^{T} \nu_{it}^2 \nu_{jt}^2 \overline{K}_{\mathbf{h}} \left( \mathbf{0} \right), \\
A_{ij,3} \left( \mathbf{x} \right) = \frac{1}{T} \sum_{t=1}^{T} e_{it} \nu_{jt} K_{\mathbf{h}} \left( \mathbf{x}_t - \mathbf{x} \right), & B_{ij,3} = \frac{1}{T} \sum_{t=1}^{T} e_{it}^2 \nu_{jt}^2 \overline{K}_{\mathbf{h}} \left( \mathbf{0} \right), \\
A_{ij,4} \left( \mathbf{x} \right) = \frac{1}{T} \sum_{t=1}^{T} e_{jt} \nu_{it} K_{\mathbf{h}} \left( \mathbf{x}_t - \mathbf{x} \right), & B_{ij,1} = \frac{1}{T} \sum_{t=1}^{T} e_{jt}^2 \nu_{it}^2 \overline{K}_{\mathbf{h}} \left( \mathbf{0} \right).
\end{array}$$

Consequently,

$$\begin{aligned}
\widehat{\Gamma} &= \sum_{i=1}^{k-1} \sum_{j=i}^{k} \int \left[ \frac{1}{T} \sum_{t=1}^{T} \left( \widehat{e}_{it} \widehat{e}_{jt} - \delta_{ij} \right) K_{\mathbf{h}} \left( \mathbf{x}_t - \mathbf{x} \right) \right]^2 d\mathbf{x} - \frac{1}{T^2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{t=1}^{T} \left( \widehat{e}_{it} \widehat{e}_{jt} - \delta_{ij} \right)^2 \overline{K}_{\mathbf{h}} \left( \mathbf{0} \right) \\
&= \sum_{i=1}^{k-1} \sum_{j=i}^{k} \left\{ \left[ \int \sum_{l=1}^{4} A_{ij,l}^2 \left( \mathbf{x} \right) + 2 A_{ij,1} \left( \mathbf{x} \right) A_{ij,2} \left( \mathbf{x} \right) + 2 A_{ij,1} \left( \mathbf{x} \right) A_{ij,3} \left( \mathbf{x} \right) + 2 A_{ij,1} \left( \mathbf{x} \right) A_{ij,4} \left( \mathbf{x} \right) \right.\right. \\
&\quad \left.\left. + 2 A_{ij,2} \left( \mathbf{x} \right) A_{ij,3} \left( \mathbf{x} \right) + 2 A_{ij,2} \left( \mathbf{x} \right) A_{ij,4} \left( \mathbf{x} \right) + 2 A_{ij,3} \left( \mathbf{x} \right) A_{ij,4} \left( \mathbf{x} \right) \right] d\mathbf{x} - \sum_{l=1}^{4} B_{ij,l} \right\}.
\end{aligned}$$

Then we can write $T \left( \mathbf{h}! \right)^{1/2} \widehat{\Gamma} = \sum_{l=1}^{10} C_{lT}$, where

$$\begin{aligned}
C_{lT} &= T \left( \mathbf{h}! \right)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \left\{ \int A_{ij,l}^2 \left( \mathbf{x} \right) d\mathbf{x} - B_{ij,l} \right\} \text{ for } l = 1, 2, 3, 4, \\
C_{lT} &= T \left( \mathbf{h}! \right)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \int A_{ij,1} \left( \mathbf{x} \right) A_{ij,l-3} \left( \mathbf{x} \right) d\mathbf{x} \text{ for } l = 5, 6, 7, \\
C_{lT} &= T \left( \mathbf{h}! \right)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \int A_{ij,2} \left( \mathbf{x} \right) A_{ij,l-5} \left( \mathbf{x} \right) d\mathbf{x} \text{ for } l = 8, 9, \text{ and} \\
C_{10T} &= T \left( \mathbf{h}! \right)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \int A_{ij,3} \left( \mathbf{x} \right) A_{ij,4} \left( \mathbf{x} \right) d\mathbf{x}.
\end{aligned}$$

The proof will be completed if we can show $C_{1T} \xrightarrow{d} N\left(0, \sigma_0^2\right)$, and $C_{lT} = o_P(1)$ for $l = 2, 3, \cdots, 10$. We only prove $C_{1T} \xrightarrow{d} N\left(0, \sigma_0^2\right)$ and $C_{lT} = o_P(1)$ for $l = 2, 3, 5$ since the other cases are similar.

We first show that $C_{1T} \xrightarrow{d} N\left(0, \sigma_0^2\right)$. Let $\boldsymbol{\varsigma}_t = (\mathbf{x}_t', \mathbf{e}_t')'$ and $\phi\left(\boldsymbol{\varsigma}_t, \boldsymbol{\varsigma}_s\right) = (\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} (e_{it}e_{jt} - \delta_{ij})\left(e_{is}e_{js} - \delta_{ij}\right) \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right)$. We can write $C_{1T} = 2T^{-1} \sum_{1 \le t < s \le T} \phi\left(\boldsymbol{\varsigma}_t, \boldsymbol{\varsigma}_s\right)$, which is a second order U-statistic and it is degenerate under the null. Under Assumptions A1, A4, and A6-A9 and the null hypothesis, one can verify the conditions of Lemma B.1 in Gao and King (2003) are satisfied so that a central limit theorem applies to $C_{1T}$. The asymptotic variance is given by $\text{plim}_{n\to\infty} 2E[\phi\left(\overline{\boldsymbol{\varsigma}}_t, \boldsymbol{\varsigma}_t\right)^2] = \sigma_0^2$, where $\overline{\boldsymbol{\varsigma}}_t$ is an independent copy of $\boldsymbol{\varsigma}_t$.

To show $C_{2T} = o_P(1)$, write

$$C_{2T} = T^{-1}(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t \neq s}^{T} \nu_{is}\nu_{js}\nu_{it}\nu_{jt} \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right)$$

By (A.11) and Assumption A5, $|\nu_{it}| = |\boldsymbol{\iota}_i'(\widehat{\mathbf{e}}_t - \mathbf{e}_t)| = |\boldsymbol{\iota}_i' \boldsymbol{\xi}_t(\overline{\theta})(\widehat{\theta} - \theta_*)| \le \overline{D}_t ||\widehat{\theta} - \theta_*||$, where recall $\boldsymbol{\xi}_t(\theta) = \partial \mathbf{e}_t(\theta)/\partial\theta'$ and $\overline{\theta}$ lies between $\widehat{\theta}$ and $\theta_*$. By Assumptions A5 and A3,

$$\begin{aligned}
|C_{2T}| &\le \frac{k(k+1)}{2} T^{-1}(\mathbf{h}!)^{1/2} \sum_{s=1}^{T} \sum_{t \neq s}^{T} \overline{D}_t^2 \overline{D}_s^2 \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right) \left\| \widehat{\theta} - \theta_* \right\|^4 \\
&= O_p(T(\mathbf{h}!)^{1/2}) O_p\left(T^{-2}\right) = o_P(1)
\end{aligned}$$

where the second line follows from a simple application of the Markov inequality, and the fact that for $t \neq s$

$$E\left[\overline{D}_t^2 \overline{D}_s^2 \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right)\right] \le \left\{E\left[\overline{D}_t^4 \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right)\right]\right\}^{1/2} \left\{E\left[\overline{D}_s^4 \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right)\right]\right\}^{1/2} = O(1).$$

Similarly, noting that $C_{3T} = T^{-1}(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t\neq s}^{T} e_{it}\nu_{jt}e_{is}\nu_{js} \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right)$, we have,

$$\begin{aligned}
|C_{3T}| &\le \frac{k(k+1)}{2} T^{-1}(\mathbf{h}!)^{1/2} \left| \sum_{s=1}^{T} \sum_{t=1}^{T} \|\mathbf{e}_t\| \|\mathbf{e}_s\| \overline{D}_t \overline{D}_s \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right) \right| \left\| \widehat{\theta} - \theta_* \right\|^2 \\
&= O_p(T(\mathbf{h}!)^{1/2}) O_p\left(T^{-1}\right) = o_P(1).
\end{aligned}$$

Noting that $C_{5T} = T^{-1}(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t=1}^{T} \left(e_{is}e_{js} - \delta_{ij}\right) \nu_{it}\nu_{jt} \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right)$, we can write $C_{5T} = C_{5T,a} + C_{5T,b}$, where

$$\begin{aligned}
C_{5T,a} &= T^{-1}(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t \neq s}^{T} \left(e_{is}e_{js} - \delta_{ij}\right) \nu_{it}\nu_{jt} \overline{K}_{\mathbf{h}}\left(\mathbf{x}_t - \mathbf{x}_s\right), \text{ and} \\
C_{5T,b} &= T^{-1}(\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{t=1}^{T} \left(e_{it}e_{jt} - \delta_{ij}\right) \nu_{it}\nu_{jt} \overline{K}_{\mathbf{h}}(\mathbf{0}).
\end{aligned}$$

By Assumptions A3, A5 and A8, and the Markov inequality,

$$
\begin{aligned}
|C_{5T,a}| &\leq \frac{k(k+1)}{2} T^{-1} (\mathbf{h}!)^{1/2} \sum_{s=1}^{T} \sum_{t \neq s}^{T} \left( \|\mathbf{e}_s\|^2 + 1 \right) \overline{D}_t^2 \overline{K}_{\mathbf{h}} (\mathbf{x}_t - \mathbf{x}_s) \left\| \widehat{\theta} - \theta_* \right\|^2 \\
&= O_p(T (\mathbf{h}!)^{1/2}) O_p \left( T^{-1} \right) = o_P (1),
\end{aligned}
$$

and

$$
\begin{aligned}
|C_{5T,b}| &\leq \frac{k(k+1)}{2} T^{-1} (\mathbf{h}!)^{1/2} \sum_{t=1}^{T} \left( \|\mathbf{e}_t\|^2 + 1 \right) \overline{D}_t^2 \overline{K}_{\mathbf{h}} (\mathbf{0}) \left\| \widehat{\theta} - \theta_* \right\|^2 \\
&= O_p((\mathbf{h}!)^{-1/2}) O_p \left( T^{-1} \right) = o_P (1),
\end{aligned}
$$

Consequently, $C_{5T} = o_P(1)$. This concludes the proof of the theorem.■

**Proof of Theorem 3.4**

Under $H_1(T^{-1/2} (\mathbf{h}!)^{-1/4})$, the expression $T (\mathbf{h}!)^{1/2} \widehat{\Gamma} = \sum_{l=1}^{10} C_{lT}$ obtained in the proof of Theorem 3.3 continues to hold. In addition, one can verify that under $H_1(T^{-1/2} (\mathbf{h}!)^{-1/4})$, $C_{lT} = o_P(1)$ continue to hold for $l = 2, 3, \cdots, 10$. The main change is associated with the term $C_{1T}$. Let $\epsilon_{ijt} = e_{it} e_{jt} - \delta_{ij}$. Let $E_t(\epsilon_{ijt})$ denote the conditional expectation of $\epsilon_{ijt}$ given $\mathcal{F}_{t-1}$ and $\overline{\epsilon}_{ijt} = \epsilon_{ijt} - E_t(\epsilon_{ijt})$. Then we can write $C_{1T} = C_{1T,a} + C_{1T,b} + C_{1T,c}$, where

$$
\begin{aligned}
C_{1T,a} &= T^{-1} (\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t \neq s}^{T} \overline{\epsilon}_{ijs} \overline{\epsilon}_{ijt} \overline{K}_{\mathbf{h}} (\mathbf{x}_t - \mathbf{x}_s), \\
C_{1T,b} &= T^{-1} (\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t \neq s}^{T} E_{\mathbf{s}} (\epsilon_{ijs}) E_t (\epsilon_{ijt}) \overline{K}_{\mathbf{h}} (\mathbf{x}_t - \mathbf{x}_s), \text{ and} \\
C_{1T,c} &= 2T^{-1} (\mathbf{h}!)^{1/2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t \neq s}^{T} \overline{\epsilon}_{ijs} E_{\mathbf{x}_t} (\epsilon_{ijt}) \overline{K}_{\mathbf{h}} (\mathbf{x}_t - \mathbf{x}_s).
\end{aligned}
$$

$C_{1T,a}$ now plays the role of $C_{1T}$ in the proof of Theorem 3.3, and we can show that $C_{1T,a} \xrightarrow{d} N\left(0, \sigma_0^2\right)$. Next, noting that under $H_1(T^{-1/2} (\mathbf{h}!)^{-1/4})$, $E_t(\epsilon_{ijt}) = T^{-1/2} (\mathbf{h}!)^{-1/4} \Delta_{ij}(\mathbf{x}_t)$, we have

$$
\begin{aligned}
C_{1T,b} &= T^{-2} \sum_{i=1}^{k-1} \sum_{j=i}^{k} \sum_{s=1}^{T} \sum_{t \neq s}^{T} \Delta_{ij} (\mathbf{x}_s) \Delta_{ij} (\mathbf{x}_t) \overline{K}_{\mathbf{h}} (\mathbf{x}_t - \mathbf{x}_s) \\
&= \frac{T-1}{T} \frac{2}{T(T-1)} \sum_{1 \leq t < s \leq T} \varphi(\mathbf{x}_t, \mathbf{x}_s) \equiv \frac{T-1}{T} \widetilde{C}_{1T,b} \qquad \text{(A.12)}
\end{aligned}
$$

where $\varphi(\mathbf{x}_t, \mathbf{x}_s) = \sum_{i=1}^{k-1} \sum_{j=i}^{k} \Delta_{ij}(\mathbf{x}_s) \Delta_{ij}(\mathbf{x}_t) \overline{K}_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}_s)$. Noticing that $C_{1T,b}$ is a second order U-statistic, a typical WLLN for U-statistic of strong mixing process (e.g., Borovkova et al., 1999) would require that $\{\varphi(\mathbf{x}_t, \mathbf{x}_s) : t, s \geq 1, t \neq s\}$ be uniformly integrable, which is difficult to verify here. By the H-decomposition, we can write

$$
\widetilde{C}_{1T,b} = \vartheta_T + 2H_T^{(1)} + H_T^{(2)}, \qquad \text{(A.13)}
$$

where $\vartheta_T = \int\int \varphi(\mathbf{x}_t, \mathbf{x}_s) f(\mathbf{x}_t) f(\mathbf{x}_s) d\mathbf{x}_t d\mathbf{x}_s$, $H_T^{(1)} = \frac{1}{T}\sum_{t=1}^{T} \varphi_1(\mathbf{x}_t) - \vartheta_T$, $H_T^{(2)} = \frac{2}{T(T-1)}\sum_{1\leq t<s\leq T}$ $\overline{\varphi}(\mathbf{x}_t, \mathbf{x}_s)$, $\varphi_1(\mathbf{x}_t) = \int \varphi(\mathbf{x}_t, \mathbf{x}_s) f(\mathbf{x}_s) d\mathbf{x}_s$, and $\overline{\varphi}(\mathbf{x}_t, \mathbf{x}_s) = \varphi(\mathbf{x}_t, \mathbf{x}_s) - \varphi_1(\mathbf{x}_t) - \varphi_1(\mathbf{x}_s) + \vartheta_T$. By the Fubini theorem and the change of variables, we have

$$\begin{aligned}
\vartheta_T &= \sum_{i=1}^{k-1}\sum_{j=i}^{k} \int\int \Delta_{ij}(\mathbf{x}_s)\Delta_{ij}(\mathbf{x}_t)\overline{K}_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}_s) f(\mathbf{x}_t) f(\mathbf{x}_s) d\mathbf{x}_t d\mathbf{x}_s \\
&= \sum_{i=1}^{k-1}\sum_{j=i}^{k} \int \Delta_{ij}^2(\mathbf{x}) f^2(\mathbf{x}) d\mathbf{x} + o(1). \qquad (A.14)
\end{aligned}$$

Note that $\varphi_1(\mathbf{x}_t)$ is a measurable function of $\mathbf{x}_t$ and inherits the $\alpha$-mixing property of the latter. By Assumption A1, $\varphi_1(\mathbf{x}_t)$ is a strictly stationary $\alpha$-mixing process with mixing coefficient $\alpha(j) \to 0$ as $j \to \infty$. By Proposition 3.44 of White (2001), $\varsigma_t$ is also ergodic. Furthermore, it is easy to verify that $E|\varphi_1(\mathbf{x}_t)| < \infty$. It follows from the Ergodic theorem (e.g., White, Theorem 3.34) that

$$H_T^{(1)} \xrightarrow{p} 0. \qquad (A.15)$$

Now, $H_T^{(2)}$ is a standard second order degenerate U-statistic with a symmetric kernel $\overline{\varphi}(\cdot,\cdot)$ : $\overline{\varphi}(\mathbf{x}_t, \mathbf{x}_s) = \overline{\varphi}(\mathbf{x}_s, \mathbf{x}_t)$ and $E\overline{\varphi}(\mathbf{x}_1, \mathbf{a}) = 0$ for any $\mathbf{a} \in \mathbb{R}^q$. Noting that

$$\max_{1<t\leq T} \max\left\{ E\left|\overline{\varphi}(\mathbf{x}_1, \mathbf{x}_t)\right|^{2(1+\delta)}, \int \left|\overline{\varphi}(\mathbf{x}_1, \mathbf{x}_t)\right|^{2(1+\delta)} dF(\mathbf{x}_1) dF(\mathbf{x}_t) \right\} = O\left((\mathbf{h}!)^{-(1+2\delta)}\right)$$

where $F(\cdot)$ is the distribution function of $\mathbf{x}_t$, it follows from Lemma C.2 of Gao and King (2003) that

$$E\left[H_T^{(2)}\right]^2 \leq C\left(\frac{2}{T(T-1)}\right)^2 T^2 (\mathbf{h}!)^{-\frac{1+2\delta}{1+\delta}} = O\left(T^{-2}(\mathbf{h}!)^{-\frac{1+2\delta}{1+\delta}}\right) = o(1).$$

Hence by the Chebyshev inequality

$$H_T^{(2)} = o_P(1). \qquad (A.16)$$

Combining (A.12)-(A.16) yields $C_{1T,b} \xrightarrow{p} \sum_{i=1}^{k-1}\sum_{j=i}^{k} \int \Delta_{ij}^2(\mathbf{x}) f^2(\mathbf{x}) d\mathbf{x} \equiv \Delta_0$.

Now, write $C_{1T,c} = C_{1T,c1} + C_{1T,c2}$, where $C_{1T,c1} = 2T^{-3/2}(\mathbf{h}!)^{1/4}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{1\leq t<s\leq T}^{T} \bar{\epsilon}_{ijs}$ $\Delta_{ij}(\mathbf{x}_t)\overline{K}_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}_s)$ and $C_{1T,c2} = 2T^{-3/2}(\mathbf{h}!)^{1/4}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{1\leq s<t\leq T}^{T} \bar{\epsilon}_{ijs}\Delta_{ij}(\mathbf{x}_t)\overline{K}_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}_s)$. By construction $E(\bar{\epsilon}_{ijs}|\mathcal{F}_{s-1}) = 0$. It follows that $E(C_{1T,c1}) = 0$ by the law of iterated expectations and the hypothesis that $(\mathbf{x}_t, \mathbf{x}_s) \in \mathcal{F}_{s-1}$ for $t < s$. By the Davydov's inequality (e.g., Bosq, 1996, p.19), we have

$$\begin{aligned}
E(C_{1T,c2}) &= 2T^{-3/2}(\mathbf{h}!)^{1/4}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{1\leq s<t\leq T}^{T} E\left[\bar{\epsilon}_{ijs}\Delta_{ij}(\mathbf{x}_t)\overline{K}_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x}_s)\right] \\
&= 2T^{-3/2}(\mathbf{h}!)^{1/4}\sum_{i=1}^{k-1}\sum_{j=i}^{k}\sum_{1\leq s<t\leq T}^{T} \int E\left[\bar{\epsilon}_{ijs}K_{\mathbf{h}}(\mathbf{x}_s - \mathbf{x})\Delta_{ij}(\mathbf{x}_t)K_{\mathbf{h}}(\mathbf{x}_t - \mathbf{x})\right] d\mathbf{x} \\
&\leq CT^{-1/2}(\mathbf{h}!)^{1/4}(\mathbf{h}!)^{-\frac{2(1+\delta)}{2+\delta}}\sum_{j=1}^{T-1}[\alpha(i)]^{\delta/(2+\delta)} = o(1) \text{ for sufficiently small } \delta > 0.
\end{aligned}$$

Similarly, we can show that $E\left(C_{1T,c1}^{2}\right)=o\left(1\right)$ and $E\left(C_{1T,c2}^{2}\right)=o\left(1\right).$ Then $C_{1T,c}=o_{P}\left(1\right)$ by the Chebyshev inequality.

Consequently, $P(\widehat{T}\geq z_{\alpha}|H_{1}(T^{-1/2}\left(\mathbf{h}!\right)^{-1/4}))\rightarrow 1-\Phi(z_{\alpha}-\Delta_{0}/\sigma_{0})$.■

# References


Alexander C. (1998), "Volatility and Correlation: Methods, Models and Applications," in *Risk Management and Analysis: Measuring and modeling Financial Risk*, John Wiley, New York, 125-172.

Alexander C. (2001), "Orthogonal GARCH," in *Mastering Risk*, Financial Times-Prentice Hall, London, 2: 21-38.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (1999), "Understanding, Optimizing, Using and Forecasting: Realized Volatility and Correlation," Working Paper, Northwestern University, Duke University and University of Pennsylvania.

Ang, A. and Chen, J. (2002), "Asymmetric Correlations of Equity Portfolios," *Journal of Financial Economics,* 63(3), 443-494.

Awartani, B. M. A. and Corradi, V. (2005), "Predicting the Volatility of the S&P-500 Index via GARCH Models: The Role of Asymmetries," *International Journal of Forecasting,* 21(1), 167-184.

Baba, Y., Engle, R. F., Kraft, D. F., and Kroner, K. F. (1991), "Multivariate Simultaneous Generalized ARCH," manuscript, Dept. of Economics, UCSD.

Bollerslev, T. (1990), "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach," *Review of Economics and Statistics,* 72, 498–505.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988), "A Capital Asset Pricing Model with Time-Varying Covariance," *Journal of Political Economy,* 96(1), 116-131.

Borovkova, S., Burton, R., and Dehling, H. (1999), "Consistency of the Takens Estimator for the Correlation Dimension", *Annals of Applied Probability* 9, 376-390.

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Springer, New York.

Cai, Z., Fan, J., and Yao, Q. (2000), "Functional-Coefficient Regression Models for Nonlinear Time Series," *Journal of American Statistical Association,* 95, 941–956.

Cai, Z., and Ould-Saïd, E. (2003), "Local M-estimator for Nonparametric Time Series," *Statistics and Probability Letters,* 65, 433-449.

Cappiello, L., Engle, R. F. and Sheppard, K. (2003), "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," ECB Working Paper No. 204.

Chen, X., Fan, Y., and Patton, A. (2004), "Simple Tests for Models of Dependence Between Multiple Financial Time Series, with Applications to U.S. Equity Returns and Exchange Rates," Manuscript.

Ding Z. (1994), *Time Series Analysis of Speculative Returns,* Ph.D. dissertation, UCSD.

Engle, R. F. (1982), "Autoregressive, Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica,* 50, 987-1007.

Engle, R. F. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business and Economic Statistics*, 20, 339-350.

Engle, R. F, and González-Rivera, G. (1991), "Semiparametric ARCH Models," *Journal of Business and Economic Statistics,* 9, 345-359

Engle, R. F. and Kroner, K. F. (1995), "Multivariate Simultaneous Generalized ARCH," *Econometric Theory* 11, 122-50.

Engle, R. F., Ng, V. R., and Rothschild, M. (1990), "Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics,* 45, 213-238.

Engle, R. F. and Sheppard, K. (2001), "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," Discussion Papers 2001-15, Dept. of Economics, UCSD.

Fama, E. and French, K. (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics,* 33, 3–56.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling*, Chapman and Hall, London.

Fan, Y. and Ullah, A. (1999), "Asymptotic Normality of a Combined Regression Estimator," *Journal of Multivariate Analysis,* 85(3), 191-240.

Gao, J., and King, M. (2003), "Estimation and Model Specification Testing in Nonparametric and Semiparametric Regression Models," *Mimeo.* Dept. of Mathematics and Statistics, Univ. of Western Australia.

Glad, I. K. (1998), "Parametrically guided non-parametric regression," *Scandinavian Journal of Statistics,* 25, 649-668.

Gozalo, P and Linton, O. (2000), "Local Nonlinear Least Squares: Using Parametric Information in Nonparametric Regression," *Journal of Econometrics,* 99, 63-106.

Hafner, C., van Dijk, D., and Franses, P. (2006), "Semiparametric Modelling of Correlation Dynamics," in T. Fomby and C. Hill (eds.) *Advances in Econometrics* 20, Part A, 59–103.

Hall, P. (1986), "On the Number of Bootstrap Simulations Needed to Construct a Confidence Interval," *Annals of Statistics,* 14, 1453-1462.

Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and Its Applications.* Academic Press, New York.

Härdle, W. and Tsybakov, A. (1997), "Local Polynomial Estimators of the Volatility Function in Nonparametric Autoregression," *Journal of Econometrics,* 81, 223-242.

Koenker, R. and Bassett, G. (1978), "Regression Quantiles," *Econometrica,* 46, 33-50.

Lee, T. H. and Long, X. D. (2008), "Copula-based Multivariate GARCH Model with Uncorrelated Dependent Errors," forthcoming in *Journal of Econometrics.*

Ledoit, O., Santa-Clara, P., and Wolf, M. (2003), "Flexible Multivariate GARCH Modeling with an Application to International Stock Markets," *Review of Economics and Statistics,* 85, 735–747.

Longin, F. and Solnik, B. (2001), "Extreme Correlation of International Equity Market," *Journal of Finance,* 56, 649-679.

Magnus, J. R. and Neudecker, H. (1999) *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley and Sons.

Mashal, R. and Zeevi, A. (2002), "Beyond Correlation: Extreme Co-movements between Financial Assets," Manuscript, Columbia University.

Mishra, S., Su, L. and Ullah, A. (2008), "Combined Estimator of Time Series Conditional Heteroskedasticity," Under review for *Journal of Business & Economic Statistics.*

Olkin, I and Spiegelman, C. (1987), "A Semiparametric Approach to Density Estimation", *Journal of American Statistical Association,* 88, 858-865.

Pagan, A. and Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press.

Pelletier, D. (2006), "Regime Switching for Dynamic Correlations," *Journal of Econometrics,* 131, 445-473.

Richardson, M. P. and Smith, T. (1993), "A Test of Multivariate Normality of Stock Returns," *Journal of Business, 66*, 295-321.

Silvennoinen, A. and Teräsvirta, T. (2005), "Multivariate Autoregressive Conditional Heteroskedasticity with Smooth Transitions in Conditional Correlations," Working Paper Series in Economics and Finance 577, Stockholm School of Economics.

Song, P., Fan, F., and Kalbfleisch, J. (2005), "Maximization by Parts in Likelihood Inference," *Journal of the American Statistical Association,* 100, 1145-1158.

Su, L. and Ullah, A. (2008), "Testing Conditional Uncorrelatedness," In press for *Journal of Business & Economic Statistics.*

Su, L., and White, H. (2007), "A Consistent Characteristic Function-based Test for Conditional Independence," *Journal of Econometrics* 141, 807-834.

Tse, Y. K. and Tsui, A. K. (2002), "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations," *Journal of Business and Economic Statistics,* 20, 351-362.

White, H. (1994), *Estimation, Inference and Specification Analysis,* Cambridge University Press.

White, H. (2001), *Asymptotic Theory for Econometricians*, 2nd ed., Academic Press.

Zangari, P. (1997), "Streamlining the Market Risk Measurement Process," *RiskMetrics Monitor,* 1, 29-35.